Evaluación y aprendizaje en educación universitaria: estrategias e instrumentos

EDITORES

Melchor Sánchez Mendiola

Adrián Martínez González





Evaluación y aprendizaje en educación universitaria: estrategias e instrumentos

Evaluación y aprendizaje en educación universitaria: estrategias e instrumentos

EDITORES

Melchor Sánchez Mendiola

Adrián Martínez González





Primera edición, junio 2022.

Evaluación y aprendizaje en educación universitaria: estrategias e instrumentos

Primera edición
UNAM, Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia, 2022

Editores
Sánchez Mendiola, Melchor
Martínez González, Adrián
Evaluación y aprendizaje en educación universitaria: estrategias e instrumentos / Sánchez Mendiola, Melchor, Martínez González, Adrián. —1ª ed. — Ciudad de México, UNAM, 2022.
p. 774
ISBN 978-607-30-6076-9
1. Educación.

La presente obra fue sometida a consideración del Comité Editorial de la Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia de la Universidad Nacional Autónoma de México, y fue aprobada por un dictamen académico a través de arbitraje por pares.

Editores
Melchor Sánchez Mendiola
Adrián Martínez González

ISBN: 978-607-30-6071-4
Impreso y hecho en México.

ÍNDICE

SECCIÓN II.

SECCIÓN III.

PRÓLOGO

El 30 de noviembre de 2015 el Rector Enrique Graue Wiechers emitió un acuerdo que reorganizó las funciones y estructura de la Secretaría General de la Universidad Nacional Autónoma de México (UNAM), creando la Coordinación de Desarrollo Educativo e Innovación Curricular (CODEIC), en la cual quedó incluida la Dirección General de Evaluación Educativa. Con esta decisión de política académica, el Rector Graue dio continuidad a los trabajos de dicha dirección general y puso énfasis en la necesidad de fortalecer la evaluación, el desarrollo educativo y la innovación curricular como ejes de una estrategia de fortalecimiento de la docencia, una de las tres funciones sustantivas de la Universidad Nacional.

A lo largo de casi cuatro años, la CODEIC llevó a cabo una importante labor que incluyó la reorganización del Consejo de Evaluación Educativa de la UNAM y la regularización de sus sesiones; el proyecto del Centro de Formación Docente que permitirá concentrar y potenciar los esfuerzos de nuestra Universidad en la formación de profesionales de la enseñanza en todos los niveles y disciplinas que se imparten en nuestra casa de estudios; la promoción de innovaciones educativas que contribuyan a modernizar y dinamizar el proceso de enseñanza-aprendizaje en los diversos niveles de estudios de la Institución, entre otros.

En junio de 2020 el Rector Graue dio un paso más en el fortalecimiento de las capacidades institucionales de la UNAM para enfrentar el desafío de la innovación educativa y consolidar nuestra oferta de educación a distancia. En el marco de la pandemia provocada por el virus COVID-19, acordó la fusión de la CODEIC con la Coordinación de Universidad Abierta y Educación a Distancia (CUAED), dando lugar a la Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia (CUAIEED). La nueva coordinación recuperó el diálogo que la CODEIC había entablado con los docentes de la Universidad, junto con la experiencia de casi medio siglo del Sistema de Universidad Abierta y Educación a Distancia, como el propósito de impulsar la innovación curricular, la evaluación de y para

el aprendizaje, así como la ampliación y plena incorporación de las modernas tecnologías de información y comunicación, del aprendizaje y el conocimiento, a todos los niveles que imparte la UNAM.

Como parte de la significativa contribución realizada por la CODEIC y que se enmarcó en la misión que le encomendó el Rector Graue, en 2020 se publicó el libro *Evaluación del y para el aprendizaje: instrumentos y estrategias*. Los cambios ocurridos en los últimos tres años, acelerados por más de dos años de pandemia y por la propia transformación de la CODEIC en CUAIEED, llevaron a los editores a plantearse el reto de llevar a cabo una profunda transformación, actualización y ampliación de ese libro, dando como resultado el nuevo volumen que el lector tiene ante sí. Constituye, como el anterior, una descripción y análisis de instrumentos y estrategias de evaluación del aprendizaje y lo que es aún más importante, para el aprendizaje y como aprendizaje, a la luz de lo aprendido en los últimos tres años, que plantearon un desafío sin precedentes a la educación. El libro pasó de 19 a 45 capítulos y se amplió el abanico de autores para incluir a especialistas de otras instituciones educativas del país y del extranjero, dando como resultado una contribución de gran calidad para abordar los temas más relevantes de la evaluación y el aprendizaje en la educación universitaria.

Se trata de un libro valioso por la calidad de sus autores y por la pertinencia de sus contribuciones, y sin duda, más completo por los temas abordados y las perspectivas que incorporaron los autores que participan en este nuevo proyecto editorial. Estos factores hacen muy recomendable su lectura para todos aquellos que estamos interesados en mejorar nuestro desempeño en el aula. Estudiar y comprender los instrumentos y estrategias de evaluación del, para y como aprendizaje, puede ayudarnos a mejorar nuestra práctica docente, que es uno de los objetivos que persigue la CUAIEED y que sin duda tendrán un impacto decisivo en la calidad de la educación y en la formación de los estudiantes. Por eso invito con entusiasmo a los profesores de la UNAM y de otras instituciones educativas a emprender la lectura este libro que sintetiza el esfuerzo y el trabajo de los destacados especialistas en los temas que aborda.

Leonardo Lomelí Vanegas
Secretario General
Universidad Nacional Autónoma de México

PREFACIO

"Evaluar es aprender."
CARLOS MAGRO MAZO

"En un bosque de cien mil árboles, no existen dos hojas iguales.
Tampoco existen dos viajes iguales en el mismo camino."
PAULO COELHO, ALEPH.

La presente obra es el resultado de una profunda transformación, actualización y ampliación de un trabajo publicado en el año 2020 por el área encargada de evaluación educativa en la Universidad Nacional Autónoma de México (UNAM), que tuvo por título *"Evaluación del y para el aprendizaje: instrumentos y estrategias"*. El libro que ahora nos ocupa se denomina *"Evaluación y aprendizaje en educación universitaria: estrategias e instrumentos"*, lo que nos lleva a recordar la frase del filósofo griego Heráclito, "no hay nada permanente, a excepción del cambio". El proceso de planeación, elaboración y publicación de la primera versión del libro ocurrió antes de la pandemia por COVID-19, por lo que se hizo necesario repensarlo en su totalidad bajo las condiciones de la nueva realidad, que afectó de forma brutal a la educación a nivel mundial, particularmente las tareas de evaluación del y para el aprendizaje. Para ello se llevaron a cabo varias reuniones con un grupo de los autores, se revisó con acuciosidad la literatura académica que se generó sobre el tema durante la pandemia, se actualizó el material, en aras de su vigencia ante los vertiginosos cambios que enfrentó el sector educativo en los tiempos recientes, y se agregaron varios temas, aspectos conceptuales, teórico-metodológicos y prácticos, incluyendo la evaluación a distancia, la evaluación sistémica, entre otros, para enriquecer la obra y proveer un panorama amplio y vigente de los aspectos relevantes que tienen que ver con la evaluación del, para y como aprendizaje en educación universitaria.

La versión anterior del libro tuvo 19 capítulos, mientras que la actual alcanza los 45, incrementando el material a más del doble. Por otra parte, se amplió el abanico de autores participantes, en esta ocasión con colaboradores de diversas organizaciones educativas y de evaluación, no solamente de México sino de otros países. El objetivo principal del libro sigue siendo ofrecer un documento de acceso abierto en español, que funcione como una herramienta útil para los docentes universitarios que requieran orientación para alinear sus prácticas de evaluación con los procesos de enseñanza y aprendizaje.

Este libro da respuesta a las necesidades que, en el ámbito de la evaluación educativa, fueron identificadas por el profesorado universitario y el área de evaluación educativa de la UNAM. Constituye la suma de conocimientos y experiencias del personal de la Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia de nuestra institución, así como de académicos y expertos en evaluación que, con iniciativa y creatividad, han desarrollado prácticas de evaluación exitosas y útiles en diversos contextos. La elaboración de este material requirió un proceso de integración de múltiples perspectivas respecto a un objeto y con una intención, pues cada uno de los capítulos que lo integran tiene como eje principal a la evaluación educativa, considerando todos aquellos ámbitos en donde esta tiene injerencia y puede ser aplicada a fin de mejorar los procesos de enseñanza y aprendizaje, así como las competencias de evaluación en los profesores de educación media superior y superior.

Las prácticas y usos de la evaluación educativa fueron puestas a prueba por la pandemia, ya que es menester documentar la adecuada formación de los estudiantes universitarios al tiempo que se propicia el aprendizaje, tomando en cuenta el momento, el contexto y el bienestar de estudiantes y docentes. Debe reconocerse que la respuesta de la comunidad del profesorado, las universidades y las organizaciones dedicadas a temas de evaluación, fue amplia, generosa y disruptiva. Nunca como ahora se habían sacudido tanto los paradigmas tradicionales de la evaluación, principalmente aquella de índole sumativa y de alto impacto, lo que ha llevado a cuestionar varias de las prácticas, usos y costumbres de los exámenes tradicionales. En este contexto, la evaluación formativa y la evaluación en línea han adquirido una gran preponderancia, pues por medio de ellas es posible contar con evidencias del grado de apropiación que tienen los estudiantes de los conocimientos, habilidades y actitudes que deben poseer al concluir un ciclo educativo.

Por otra parte, las exigencias normativas e institucionales de evaluaciones finales, calificaciones y resultados cuantitativos para la toma de decisiones de admisión o promoción, plantea una gran cantidad de retos técnicos, prácticos y éticos, que están enfrentándose de forma diversa y asertiva en las aulas, escuelas y universidades a nivel global. Por todo ello, los saberes y habilidades que han de poseer los docentes-evaluadores respecto a las diversas concepciones teórico-metodológicas que están presentes en el campo de la evaluación educativa, requieren estas características fundamentales: contar con evidencias de su utilidad, pertinencia y aplicabilidad, así como ser flexibles para responder a las necesidades específicas de cada contexto educativo.

Esperamos que los materiales desarrollados en este libro aporten información y conocimientos útiles, relevantes y prácticos para todos los docentes que deseen enriquecer su formación como evaluadores. Sabemos que la evaluación es un proceso inherente al ejercicio docente que es complejo, no tan sencillo de aplicar, y que puede estar salpicado de incertidumbre y frustración, pero estamos convencidos que si se planea y aplica con el propósito de mejorar y consolidar el aprendizaje de los estudiantes, tiene enorme potencial para participar con el estudiantado en el hermoso viaje del aprendizaje. Es importante que las y los profesores cuenten con una formación adecuada para llevar a cabo la evaluación y su

integración con la enseñanza de manera exitosa. Invitamos al profesorado universitario a continuar compartiendo y poniendo en práctica todos los saberes y habilidades que han adquirido a lo largo de su desarrollo profesional, reflexiones y experiencias, al tiempo que incorporan los conceptos vertidos en este libro en sus tareas educativas.

Dr. Melchor Sánchez Mendiola
Coordinador de Universidad Abierta, Innovación Educativa y Educación a Distancia.
Profesor Titular C de Tiempo Completo, División de Estudios de Posgrado, Facultad de Medicina.
Universidad Nacional Autónoma de México. Ciudad de México, México

Dr. Adrián Martínez González
Director de Evaluación Educativa, Coordinación de Universidad Abierta,
Innovación Educativa y Educación a Distancia.
Profesor de Carrera Titular C de Tiempo Completo, Facultad de Medicina.
Universidad Nacional Autónoma de México. Ciudad de México, México

Ciudad de México,
junio de 2022.

Sección I.

ASPECTOS CONCEPTUALES Y METODOLÓGICOS

Capítulo 1

EVALUACIÓN DEL, PARA Y COMO APRENDIZAJE

Melchor Sánchez Mendiola

"Evaluación es un intento de conocer a la persona."
Derek Rowntree

"No todo lo que puede ser contado cuenta,
y no todo lo que cuenta puede ser contado."
Albert Einstein

INTRODUCCIÓN

La palabra evaluación genera múltiples reacciones en las personas, no todas agradables. Solemos decir que las actividades organizacionales deben ser evaluadas, que para mejorar hay que evaluar, que sin evaluación es imposible saber exactamente en dónde estamos, entre otras sentencias, pero cuando se trata de que la evaluación sea personal la situación cambia, sobre todo cuando es sumativa de alto impacto. Al tener noticia de que seremos evaluados, frecuentemente nos genera una sensación de incomodidad, de angustia o hasta miedo, y si podemos diferirla o exentarla en ocasiones, lo hacemos. Evaluar se dice fácil, pero en los hechos no lo es tanto, requiere una actitud reflexiva y madura, recursos para realizarla, personal con entrenamiento y experiencia en sus matices metodológicos y técnicos, tiempo para planearla, realizarla y analizarla, así como infraestructura para documentar las diferentes etapas del proceso. Los procesos de evaluación efectivos requieren estructuras organizacionales participativas, no tan verticales o jerárquicas, que estén dispuestas a aceptar los resultados con entusiasmo y transparencia, para actuar en consecuencia y mejorar la estructura, procesos y resultados del sistema. Se requiere pensamiento sistémico y visión de largo plazo para que el proceso de evaluación se integre adecuadamente al sistema, y participación activa de las personas que conforman los diferentes elementos del mismo. En resumen, la evaluación no es una tarea fácil ni sencilla, se requiere esfuerzo a nivel individual y colectivo, así como apoyo de los diversos niveles de la estructura organizacional.

Reflexionemos sobre los siguientes escenarios educativos:

- Un profesor de matemáticas en la escuela de arquitectura hace un pequeño examen de tres preguntas al inicio de cada clase, sobre el tema de la misma. Al final del curso

promedia los resultados de estas pruebas y los incluye como 5% de la calificación final.

- El departamento de Biología Molecular y Genómica de una escuela de medicina decide, por acuerdo del profesorado, que los exámenes departamentales elaborados por su cuerpo colegiado sean el principal elemento de la calificación final, siendo el 70% de la misma.
- Una docente de la clase de filosofía de la educación en la carrera de pedagogía, rechaza el uso de los exámenes escritos en sus cursos, solamente usa exámenes orales porque son más realistas y permiten explorar pensamiento crítico.
- Un aspirante a ingresar a la universidad, que tiene niveles altos de empatía, habilidades de comunicación, creatividad y curiosidad, no alcanza la puntuación requerida en el examen de conocimientos. Como consecuencia presenta un cuadro importante de depresión.
- Un profesor de cinematografía utiliza gran parte de su tiempo en reuniones individuales con los alumnos, en esos espacios les da retroalimentación detallada y constructiva. Ni los estudiantes ni él identifican estos eventos como actos de evaluación.
- Una profesora de ciencias políticas en la universidad, acostumbrada a usar reactivos de falso y verdadero en sus exámenes, se niega a dejar de utilizarlos a pesar de la abundante evidencia publicada que documenta limitaciones de este formato de preguntas.
- Un estudiante de ingeniería con promedio de 8.5 en el curso, le exige al profesor que redondee la calificación a 9 ya que no sería justo que le disminuyan la puntuación a 8, lo que afectaría su promedio en la carrera.
- El departamento académico de una universidad compara los resultados de los exámenes finales de los cursos de los últimos cinco años, sin tomar en cuenta que los exámenes han sido de diferentes grados de dificultad en cada ciclo.

Todas y cada una de estas viñetas son realistas y tienen diversas implicaciones educativas. En cada una de ellas hay diferentes actores que tienen una serie de experiencias y sesgos hacia la evaluación, que determinan su actitud ante estas situaciones. Estos escenarios tienen implicaciones técnicas, pedagógicas, didácticas, psicológicas, afectivas y socioemocionales, que deben ser analizadas y atendidas con profesionalismo y niveles apropiados de conocimientos sobre el proceso evaluativo. La evaluación como campo de estudio ha crecido y se ha hecho más sofisticada en el último siglo, con incremento sustancial en trabajos de investigación, publicaciones, programas de posgrado, agrupaciones académicas especializadas, y congresos sobre el tema. Sin embargo, han ocurrido una serie de vaivenes en la actitud de la sociedad, las universidades, el profesorado y el estudiantado hacia este campo disciplinario, que con frecuencia generan confusión y hasta rechazo. Es necesario que el profesorado que participa en educación a todos los niveles tenga una comprensión operativa básica de los conceptos relevantes y pertinentes de evaluación en educación, para poder lograr cabalmente el cumplimiento de la labor de la enseñanza.

En este libro nos enfocamos en el uso de la evaluación y sus implicaciones para el aprendizaje, en virtud de que consideramos que los logros educativos de los educandos son de primordial importancia, y del hecho que es indispensable no solo documentar el aprendizaje, sino apoyarlo y propiciarlo. La enseñanza debe integrarse con los mecanismos e

instrumentos de evaluación disponibles para lograr sus metas. Recordemos los comentarios de Sigmund Freud sobre las *profesiones imposibles*, aquellas en las que puedes estar seguro de lograr resultados insatisfactorios, como son el psicoanálisis, el acto de gobernar y la educación (Freud, 1937). Los docentes creemos que lo que enseñamos es aprendido por los estudiantes, desafortunadamente esto no es así, siempre hay un gradiente entre lo que se enseña y lo que se aprende, y también se aprenden cosas diferentes de lo que se pretende enseñar. Por ello, la única manera de tener cierta claridad sobre el efecto de la educación universitaria y su impacto en los estudiantes, es llevar a cabo una evaluación técnicamente adecuada, alineada con el currículo y los métodos de enseñanza, que proporcione resultados interpretables y útiles para los diferentes actores del proceso educativo (Brennan, 2006; Martone y Sireci, 2009). En este capítulo describiremos algunas de las definiciones relevantes a la evaluación del, para y como aprendizaje, y sus implicaciones educativas.

¿QUÉ ES EVALUACIÓN?

Hay muchas definiciones del término *evaluación* en el área educativa, Miller la define como: "un término genérico que incluye un rango de procedimientos para adquirir información sobre el aprendizaje del estudiante y la formación de juicios de valor…" (Miller, 2012). Ello implica un proceso sistemático de acopio de información mediante la aplicación de diversos instrumentos, como son exámenes escritos u orales, para ser analizada con rigor metodológico y así fundamentar la toma de decisiones. La edición más reciente de los Estándares para Pruebas Educativas y Psicológicas de AERA-APA-NCME, define evaluación como: "método sistemático de obtención de información, usado para formular deducciones sobre las características de personas, objetos o programas; proceso sistemático para medir o evaluar las características o el desempeño de individuos, programas u otras entidades con la finalidad de hacer inferencias; en ocasiones se usa como sinónimo de prueba" (AERA, APA y NCME, 2014).

Una prueba (examen o test) es un tipo particular de evaluación que consiste en un conjunto de preguntas o retos, administrados en un periodo de tiempo fijo y bajo condiciones razonablemente comparables para el estudiantado (Miller, 2012). De acuerdo a los Estándares de AERA-APA-NCME una prueba es un "dispositivo de evaluación o procedimiento en el cual se obtiene y puntúa una muestra sistemática del comportamiento de un examinado en un dominio específico, a través de un proceso estandarizado" (AERA, APA y NCME, 2014). Las pruebas son uno de los instrumentos utilizados primordialmente para evaluación sumativa y diagnóstica.

Es pertinente enfatizar la diferencia entre "evaluación" y "medición", ya que medición es la asignación una descripción numérica a los resultados de una prueba u otro tipo de evaluación de acuerdo a reglas específicas. En ese sentido el enfoque de una medición es más limitado y cuantitativo, mientras que evaluación se conceptualiza como un término más amplio e incluyente, que puede incluir elementos tanto cuantitativos como cualitativos, e incluye juicios de valor sobre el proceso y los resultados (Miller, 2012).

Por otra parte, es importante comentar el uso de las palabras en otros idiomas, ya que en español usamos la misma palabra para evaluación del aprendizaje y evaluación de programas,

mientras que en inglés el término más frecuentemente usado para evaluar aprendizaje es "*assessment*", y para evaluar programas es "*evaluation*". Esto puede prestarse a confusión al usar la literatura académica internacional, la cual predominantemente está en inglés (aunque esto no es una regla, ya que varía dependiendo del contexto y del país) (McKean y Aitken, 2016). La terminología utilizada en educación y particularmente en evaluación educativa es frecuente fuente de discusión, desacuerdos y confusión, por lo que es importante mantener una actitud abierta y reflexiva a los diversos usos de los términos en diferentes países y grupos profesionales, así como consultar los glosarios internacionales de estos términos, como el de la UNESCO https://learningportal.iiep.unesco.org/en/glossary.

Independientemente de las definiciones técnicas que usemos de evaluación y sus conceptos cercanos, el profesorado que tiene interacciones con estudiantes debe internalizar la evaluación desde una visión más profunda, como sugirió en 1977 Derek Rowntree, de Australia: "Cuando una persona, con algún tipo de interacción directa o indirecta con otra, obtiene e interpreta información de manera consciente sobre el conocimiento y la comprensión, habilidades y actitudes de la otra persona. Hasta cierto punto *evaluación es un intento de conocer a esa persona*" (Rowntree, 1977). Se deben tener en cuenta los siguientes principios generales en evaluación educativa (Miller, 2012):

1) Es prioritario especificar claramente lo que se va a evaluar.
2) La evaluación es un medio para un fin, no un fin en sí mismo.
3) Los métodos de evaluación deben elegirse con base en su relevancia, tomando en cuenta las características que se van a evaluar.
4) Se requiere una variedad de procedimientos e instrumentos para que la evaluación sea útil y efectiva.
5) Su uso adecuado requiere conocimiento sobre las bases técnicas de la evaluación y sus instrumentos, así como sobre las bondades y limitaciones de cada método.

En este libro nos enfocamos en los aspectos de la evaluación que tienen implicaciones en el aprendizaje, por lo que describiremos los conceptos de evaluación **del** aprendizaje, **para** el aprendizaje y **como** aprendizaje. Como será aparente para el lector, el uso de estas preposiciones y mediadores terminológicos son de particular importancia en la actualidad, cuando la evaluación debe convertirse en un elemento consustancial de la enseñanza y el aprendizaje. Debemos potenciar el uso de los avances logrados en este campo para que los estudiantes y docentes no sufran la evaluación, que no la vivan como una serie interminable de exámenes que generan estrés, ansiedad y desgaste, para en su lugar lograr un aprendizaje más auténtico e integrado (Bennett, 2015; Brennan, 2006; Moreno-Olivos, 2010). Al final del día, la evaluación y el aprendizaje son partes de un todo que puede visualizarse como un ciclo (Figura 1).

En el esquema del ciclo pueden apreciarse diversos elementos, todos fundamentales para que el proceso se convierta en un círculo virtuoso: la misión de la institución educativa con sus metas y objetivos de aprendizaje, la identificación de los constructos a evaluar, la

colección de evidencia de aprendizaje durante el proceso, la interpretación de dicha evidencia y su uso para retroalimentación y mejora continua, y la implementación reflexiva de cambios en el proceso educativo, para aplicarse en subsecuentes ciclos. Los conceptos descritos a continuación son analizados con mayor profundidad en otros capítulos del libro, se presentan aquí como parte del panorama general del proceso de evaluación y sus implicaciones para el aprendizaje.

Figura 1. El ciclo de la evaluación y el aprendizaje (adaptado de Maki, 2010)



CLASIFICACIÓN TRADICIONAL DE TIPOS DE EVALUACIÓN

Evaluación diagnóstica, sumativa y formativa. Una de las clasificaciones tradicionales de la evaluación educativa, desde el punto de vista de su objetivo, es: diagnóstica, sumativa y formativa.

La evaluación *diagnóstica* se realiza al principio de un curso o actividad académica, con la finalidad de determinar el nivel de conocimiento, habilidad o actitud del estudiante (Capítulo 3). Esta información puede ser de utilidad para las instituciones educativas y el profesorado, ya que permite hacer adecuaciones en el contenido y en la implementación de las actividades programadas que correspondan a las características del alumnado.

La evaluación *sumativa* es aquella compuesta por la suma de valoraciones efectuadas durante un curso o unidad didáctica, a fin de determinar el grado con que los objetivos de la instrucción se alcanzaron, otorgar calificaciones o certificar competencias (Capítulo 5). Este tipo de evaluación se inserta en el marco de "evaluación **del** aprendizaje", que se comentará

más adelante en este capítulo. Un tipo de exámenes sumativos merece atención especial, los llamados *exámenes de altas consecuencias o de alto impacto* (Capítulo 5) (Sánchez y Delgado, 2017).

La evaluación *formativa* es la que se utiliza para monitorizar el progreso del aprendizaje con la finalidad de proporcionar retroalimentación al estudiante sobre sus logros, deficiencias y oportunidades de mejora (Capítulo 4). Esta evaluación ocurre a lo largo de todo el proceso educativo del estudiantado, puede ser *formal* o *informal*. La evaluación formativa tiene un poderoso componente educativo, ya que permite identificar aquellas actividades que se llevan a cabo correctamente para continuar realizándolas de dicha manera, y aquellas que poseen alguna deficiencia, poder detectarlas a tiempo y corregirlas (Man Sze Lau, 2016; Martínez Rizo, 2009). Este tipo de evaluación forma parte de la *evaluación **para** el aprendizaje*, el cual se comentará más adelante.

EVALUACIÓN REFERIDA A NORMA Y CRITERIO

Otra forma tradicional de clasificar a la evaluación es de acuerdo con la interpretación de los resultados. Puede ser con referencia a norma (relativa), o con referencia a criterio (absoluta). Cuando los resultados de la evaluación se interpretan con *referencia a norma*, el resultado se describe en términos del desempeño del grupo y de la posición relativa de cada uno de los estudiantes evaluados (Miller, 2012; Montgomery y Connolly, 1987). Este tipo de evaluación se utiliza para colocar a los alumnos en escalas de rendimiento y puntaje, con la finalidad de asignarles un lugar dentro del grupo. Ejemplos de este tipo de interpretación son: la cantidad de errores de ortografía en un manuscrito es igual o mayor que la del 60% de su grupo; la puntuación en el examen de biología del estudiante está en el percentil 75 de una muestra nacional que ingresó a la universidad.

La evaluación con *referencia a criterio* describe el resultado específico que se encontró, de acuerdo a criterios o metas preestablecidos. Este tipo de evaluación busca la comparación del estudiante con relación a un estándar establecido previamente. Ejemplos de este tipo de evaluación son: los exámenes de inglés como segunda lengua (*Test of English as a Foreign Language*, TOEFL), en el que hay niveles de desempeño previamente determinados y los resultados se interpretan de acuerdo a dichos estándares, no de acuerdo al desempeño del grupo de sustentantes; un examen de certificación de una especialidad médica, en el que se establece un nivel determinado de desempeño específico para acreditarla. En el contexto de la evaluación criterial, puede darse el caso de que ningún estudiante apruebe el examen al no alcanzar el mínimo establecido, lo que no ocurre con las evaluaciones con referencia a norma o relativas.

La evaluación normativa versus la evaluación criterial tiene diversas aristas (Lok et al, 2016). Mientras que los resultados de un examen criterial pueden utilizarse para jerarquizar a los estudiantes por la puntuación obtenida, en un examen normativo también pueden definirse niveles de desempeño. Desde un punto de vista estricto, la interpretación con referencia a norma o a criterio se refiere solo al método de interpretación de los resultados, no tanto al examen en sí, de manera que estas interpretaciones serán más significativas y útiles

en tanto que las evaluaciones se diseñen específicamente para la interpretación a realizar. Los dos tipos de evaluación descritos se pueden conceptualizar como extremos de un continuo, más que una dicotomía estricta, en el que el extremo de la evaluación con referencia a criterio se enfoca en la descripción del desempeño de la persona, y el extremo con referencia a norma enfatiza la discriminación o identificación de los individuos (Miller, 2012). En el Tabla 1 se describen las principales diferencias entre estos tipos de evaluación.

Tabla 1. Comparación de las características de la evaluación con referencia a criterio y con referencia a norma (adaptado de Lok et al, 2016; Miller, 2012; Montgomery y Connolly, 1987)

	Evaluación con referencia a criterio (absoluta)	Evaluación con referencia a norma (relativa)
Propósito	• Refleja el progreso del desarrollo de estudiantes individuales. • Usada para planeación de programas y certificación de individuos.	• Genera un ordenamiento de los estudiantes por sus posiciones relativas en el grupo • Utilizada para diagnóstico y colocación de estudiantes.
Diseño de tareas evaluadas	Alineadas con el contenido y resultados de aprendizaje esperados.	Discrimina a los estudiantes de alto y de bajo desempeño.
Unidad de interpretación de las puntuaciones	Individual.	Grupal.
Presentación de los resultados	Calificaciones ligadas a criterios determinados.	Calificaciones derivadas de puntuaciones crudas, generalmente presentadas en una campana de distribución o en niveles gruesos de puntuaciones.
Ventajas	• Capacidad de reflejar el desempeño real y el progreso de cada estudiante. • Permite la incorporación explícita de habilidades cognitivas de alto nivel. • Es más defendible si se establecen adecuadamente los estándares de desempeño.	• Permite comparar a los individuos dentro del grupo. • Puede ajustarse estadísticamente para tener una amplitud específica (por ejemplo: desviación estándar). • Puede evitar la "inflación de calificaciones".
Desventajas	• No hay control en la distribución de las calificaciones, puede llevar a "inflación de calificaciones". • Limitada capacidad de comparar el desempeño individual contra el del grupo. • Puede enfrentar obstáculos para su adopción, si la organización desconoce la metodología y está acostumbrada a la evaluación con referencia a norma.	• El desempeño del grupo puede influir las calificaciones individuales, lo que puede ser injusto. • Puede no mostrar el rango de diferencias reales entre los estudiantes individuales • No nos dice con precisión si los individuos son competentes o tienen las habilidades requeridas para una tarea.

INSTRUMENTOS DE EVALUACIÓN DEL APRENDIZAJE

Un aspecto indispensable de la evaluación educativa son las herramientas o instrumentos que utiliza. Los instrumentos de evaluación son técnicas de medición y recolección de datos que tienen distintos formatos, atendiendo a la naturaleza de la evaluación. Existe una gran variedad de instrumentos con diversas ventajas y limitaciones para documentar el aprendizaje de los conocimientos, habilidades y destrezas de los estudiantes. Es responsabilidad del profesor y de las instituciones educativas elegir los métodos apropiados para el proceso de evaluación. Estos pueden clasificarse en las siguientes categorías:

- *Evaluaciones escritas*: ensayos, preguntas directas de respuesta corta, exámenes de opción múltiple, relación de columnas, disertaciones, reportes.
- *Evaluaciones prácticas*: exámenes orales, exámenes prácticos con casos, examen clínico objetivo estructurado (ECOE).
- *Observación*: reporte del profesor, listas de cotejo, rúbricas.
- *Portafolios y otros registros del desempeño*: libretas de registro, portafolios, registros de procedimientos.
- *Autoevaluación y evaluación por pares*: reporte del educando, reporte de los compañeros.

Cada uno de estos instrumentos tiene ventajas y desventajas, características psicométricas, así como recomendaciones para su implementación. Es responsabilidad de los profesores y responsables de la evaluación en las instituciones educativas diseñar, seleccionar, implementar y acumular evidencia de validez de los instrumentos más apropiados para evaluar el aprendizaje de los estudiantes, de acuerdo al currículo y las características del contexto local.

CRITERIOS PARA UNA BUENA EVALUACIÓN

La evaluación educativa es tan buena como la metodología utilizada y el uso que se hace de los resultados. Varias organizaciones internacionales han propuesto criterios sobre las *buenas prácticas* en evaluación, a continuación, describimos brevemente algunos (Norcini et al, 2011):

1) **Validez.** Uno de los conceptos más importantes para que los resultados de los procesos de evaluación tengan sustento sólido y uso apropiado, es el de *validez*. El capítulo 2 de este libro explora este tema con mayor profundidad.
2) **Confiabilidad o fiabilidad**. Es la capacidad del examen de arrojar un resultado consistente cuando se repite, es decir, la reproducibilidad del examen. El concepto también es analizado en el capítulo 2.
3) **Justicia y equidad**. En las últimas décadas las principales organizaciones de evaluación educativa del mundo han enfatizado la necesidad de justicia y equidad en todo el proceso educativo, incluyendo la evaluación del aprendizaje, para ser congruentes con el principio social de la educación (AERA, APA y NCME, 2014).

4) **Equivalencia**. Se refiere a que los exámenes proporcionen puntuaciones o decisiones equivalentes, cuando se administran en diferentes instituciones o tiempos (AERA, APA y NCME, 2014; Norcini et al, 2011). Es un concepto poco conocido por el profesorado, a pesar de su importancia para interpretar exámenes aplicados de manera periódica que pretenden evaluar el mismo constructo, o exámenes en diferentes contextos en los que queremos asegurar que sean de la misma dificultad, sobre todo en evaluación sumativa de alto impacto (Carter, 1984). Para lograr equivalencia se requieren procedimientos estadísticos de varios grados de sofisticación, que se encuentran en el grupo de métodos de equiparación o *igualación* de exámenes. Uno de ellos es el uso de *reactivos ancla* (preguntas con un grado de dificultad similar y comportamiento estadístico bien documentado) en un porcentaje de reactivos de cada versión del examen. Para el uso óptimo de estas técnicas se requieren profesionales de evaluación educativa expertos en estos procedimientos (capítulo 20).
5) **Factibilidad y aceptabilidad**. Estas propiedades se refieren a que las evaluaciones sean prácticas, realistas y apropiadas a las circunstancias del contexto, incluyendo instalaciones físicas, recursos humanos y financieros. Por ejemplo, el método más utilizado en el mundo para evaluar la competencia clínica en medicina es el Examen Clínico Objetivo Estructurado, que consiste en una serie de múltiples estaciones estandarizadas, en las que el sustentante se enfrenta en cada una a un reto que requiere la aplicación de competencias específicas, como pueden ser habilidades de comunicación, realizar un diagnóstico, o interpretar estudios de laboratorio (Boursicot et al., 2011). Este tipo de examen requiere gran cantidad de recursos humanos, instalaciones apropiadas y mucha dedicación en términos de disciplina, tiempo y organización. Esta disponibilidad de recursos puede no estar al alcance de varias escuelas, de manera que, aunque el examen tenga abundante evidencia de validez y confiabilidad, si no es factible deben buscarse otras alternativas como los exámenes orales o la observación directa. Otros ejemplos podrían ser el uso de exámenes adaptativos por computadora, o simuladores de alta fidelidad, herramientas que requieren gran inversión inicial y de mantenimiento. Las evaluaciones deben ser aceptables tanto para los estudiantes como para los profesores. Si hay un rechazo de la comunidad a algún tipo de evaluación, por ejemplo, a la evaluación por pares, eso hará difícil su implementación.
6) **Efectos educativo y catalítico.** Los métodos de evaluación, sobre todo los de índole sumativa, tienen efectos en los métodos de estudio y las prioridades de aprendizaje de los estudiantes (Newble, 1983). Aunque el profesor le diga al estudiante que un tema es fundamental, la pregunta inevitable es: "¿eso va a venir en el examen?" La cultura prevalente en varios sitios es que si algo no cuenta para el examen no se le da mucha importancia, por lo que la manera como se aplica la evaluación tiene efectos en la motivación de los estudiantes y en sus métodos de estudio. La evaluación tiene también un efecto *catalítico*, ya que puede tener influencia en los demás docentes, los departamentos académicos y la misma institución (Norcini et al., 2011). Si se privilegian los exámenes escritos de opción múltiple, habrá un efecto en cascada en los diferentes actores del proceso educativo. Si

se fomenta la evaluación formativa, de la misma manera pueden cambiar las actitudes hacia la evaluación de los participantes, sobre todo cuando viven sus efectos positivos.

¿QUÉ ES LA EVALUACIÓN DEL, PARA Y COMO APRENDIZAJE?

En las últimas décadas ha ocurrido un movimiento pendular del énfasis en la evaluación sumativa de alto impacto, con exámenes escritos de opción múltiple, estandarizados a gran escala, hacia una evaluación más amplia y profunda, con una visión más holística de su rol en el proceso de enseñanza-aprendizaje (Bennett, 2015; Hodges, 2013). El paradigma preponderante desde mediados del siglo XX, dominado por el discurso de la psicometría y la objetividad, en el que las conductas y tareas humanas son convertidas en números, está siendo modificado por el movimiento hacia un uso más adecuado de la evaluación como parte del proceso de aprendizaje. Las ventajas tradicionales de los exámenes estandarizados objetivos, de aparente justicia y equidad con igualdad de condiciones para los participantes, han sido fuertemente cuestionadas y el concepto de evaluación **para** el aprendizaje ha adquirido fuerza.

La importancia de evaluaciones más auténticas, en ambientes de trabajo más dinámicos e interprofesionales que requieren trabajo en equipo, han impulsado un importante cambio en las prioridades de la sociedad, las universidades, el profesorado, el mercado laboral y el estudiantado. Las evaluaciones sumativas de alto impacto requieren gran inversión en recursos financieros, humanos, técnicos y de tiempo, que si bien siguen teniendo importancia en contextos específicos, deben balancearse con otras estrategias y visiones de la evaluación que ayuden a los estudiantes a aprender, que rescaten los aspectos éticos, socioemocionales y humanos de la educación, usando metodologías cualitativas, mixtas, que exploten la riqueza de la subjetividad que ocurre en los espacios de aprendizaje formales e informales. En diversos capítulos del libro se abordan temas relativos a ello (capítulos 35, 37, 38, 40). Por otra parte, la pandemia por COVID-19 ha generado un verdadero tsunami educativo a nivel global, con un fuerte impacto en los exámenes sumativos en ambientes controlados, con una migración a las evaluaciones en línea, incluso en casa de los estudiantes. Esto tiene diversas implicaciones que se comentan en los capítulos correspondientes (capítulos 5 y 9).

En los últimos años ha adquirido fuerza el concepto de "evaluación **del-para-como** aprendizaje" ("*assessment* ***of-for-as*** *learning*" en inglés), que pretende modificar el énfasis que ha existido en la evaluación sumativa, exámenes y calificaciones, hacia un panorama más amplio e integrado que nos lleve a anclar todo el proceso de evaluación con el aprendizaje, la meta fundamental del proceso educativo (Ashford-Rowe et al, 2014; Bennett, 2015; Harapnuik, 2021; Maki, 2010; NFETLHE, 2017a). Si bien en español se escucha un poco extraño combinar las palabras del, para y como, con evaluación y aprendizaje, sobre todo cuando se verbalizan oralmente, creemos que el argumento más importante para apoyar el uso y aplicación de este concepto en nuestro medio, es centrar los esfuerzos en el aprendizaje. Propiciar que la evaluación no sea un elemento con cierto carácter punitivo que empuja o determina el aprendizaje, sin tomar en cuenta la libertad de cátedra, el currículo formal

y oculto, el bienestar de los estudiantes, para poder migrar a un panorama más integrado y holístico. En última instancia el propósito principal de la evaluación es potenciar y mejorar el aprendizaje, no solo documentarlo, por lo que el aprendizaje idealmente debe ser el centro de todas las evaluaciones (McKean y Aitken, 2016). La enseñanza, el aprendizaje y la evaluación son conceptos y actividades que están ligados de manera inextricable, y el alineamiento entre estos elementos con la planeación, diseño e implementación curricular es indispensable y se convierte en un elemento clave para el éxito educativo. En el siguiente apartado describiremos por separado los tres elementos "del", "para" y "como", para después integrarlos en un modelo que toma lo mejor de los tres y lo liga con el resto del proceso educativo.

- **Evaluación DEL aprendizaje** ("*assessment OF learning*"). De acuerdo con varios autores este tipo de evaluación es equivalente a la evaluación sumativa (capítulo 5), para documentar que el aprendizaje ocurrió y el nivel del mismo. Su naturaleza es evaluar actividades que ya ocurrieron, después o al final de un periodo de aprendizaje, y enfatiza los aspectos cuantitativos y numéricos, asociándose con las calificaciones o grados. Cuando esta evaluación tiene consecuencias importantes en el estudiante, se le denomina "evaluación de alto impacto" (capítulo 5). En este tipo de evaluación el actor principal es el profesor o la organización que aplica la evaluación, quienes son los principales tomadores de decisiones, y el estudiantado es un participante pasivo que recibe o a quien se le aplica el examen o la prueba, en contraste con la evaluación para el aprendizaje (Figura 2 y Tabla 2).
- **Evaluación PARA el aprendizaje** ("*assessment FOR learning*"). Como se comentó anteriormente, la principal meta de la evaluación debe ser mejorar el aprendizaje, no solo medirlo, por lo que cuando hablamos de evaluación para el aprendizaje nos referimos a la evaluación tradicionalmente llamada formativa, ligada a la retroalimentación (Maki, 2010; Man Sze Lau, 2016; Martínez Rizo, 2009; Wiliam, 2011). Esta evaluación ocurre durante todo el proceso de enseñanza y aprendizaje, es más longitudinal, representa un diálogo que ocurre entre docentes y estudiantes a lo largo de sus múltiples interacciones. Está enfocada en ayudar al estudiante, identificar sus áreas de oportunidad y logros, para orientarlo a progresar de una mejor manera en el proceso educativo, sin generarle estrés o desgaste, tratándolo como persona. Pretende moverse de una acción que se le hace **al** estudiante, a un proceso que se hace **con** el estudiante. Esta evaluación es inseparable de la enseñanza y apoya fuertemente el aprendizaje, si se lleva a cabo con profesionalismo y responsabilidad (Figura 2, Tabla 2).
- **Evaluación COMO aprendizaje** ("*assessment AS learning*"). En este tipo de evaluación el estudiante es empoderado, tiene mayor responsabilidad en el proceso de aprendizaje y puede ser el tomador clave de decisiones. El estudiantado necesita adquirir habilidades para el uso de los conceptos básicos de evaluación en su desarrollo personal. El aprendizaje autodirigido a lo largo de la vida, el aprendizaje autónomo, el pensamiento crítico, entre otros, requieren evaluar datos e información sobre situaciones laborales y vivenciales, analizarlos, establecer juicios de valor, tomar decisiones sobre la problemática

personal y profesional. Todo ello requiere autoevaluación y capacidad de tomar decisiones con base en evaluación de contextos y realidades complejas. Aunque el profesorado generalmente es quien tiene el poder jerárquico en el proceso educativo formal, la evaluación como aprendizaje mueve este locus de control externo a un control más intrínseco por parte del estudiantado. Sin embargo, el estudiante requiere apoyo de docentes y sus pares para ejercer cabalmente las habilidades mencionadas. La evaluación como aprendizaje ayuda a los estudiantes a aprender a aprender, fomenta la meta-cognición y el aprendizaje autorregulado (Figura 2, Tabla 2).

Diversos autores también incluyen en el paraguas conceptual de evaluación como aprendizaje, al aprendizaje potenciado por los exámenes (*test-enhanced learning*), uno de los beneficios importantes de la evaluación para mejorar el aprendizaje a través de pruebas, descrito en el capítulo 7 (Green et al, 2018; Roediger et al, 2011).

Figura 2. Elementos de la evaluación del, para y como aprendizaje (adaptado de National Forum for the Enhancement of Teaching and Learning in Higher Education, 2017a)

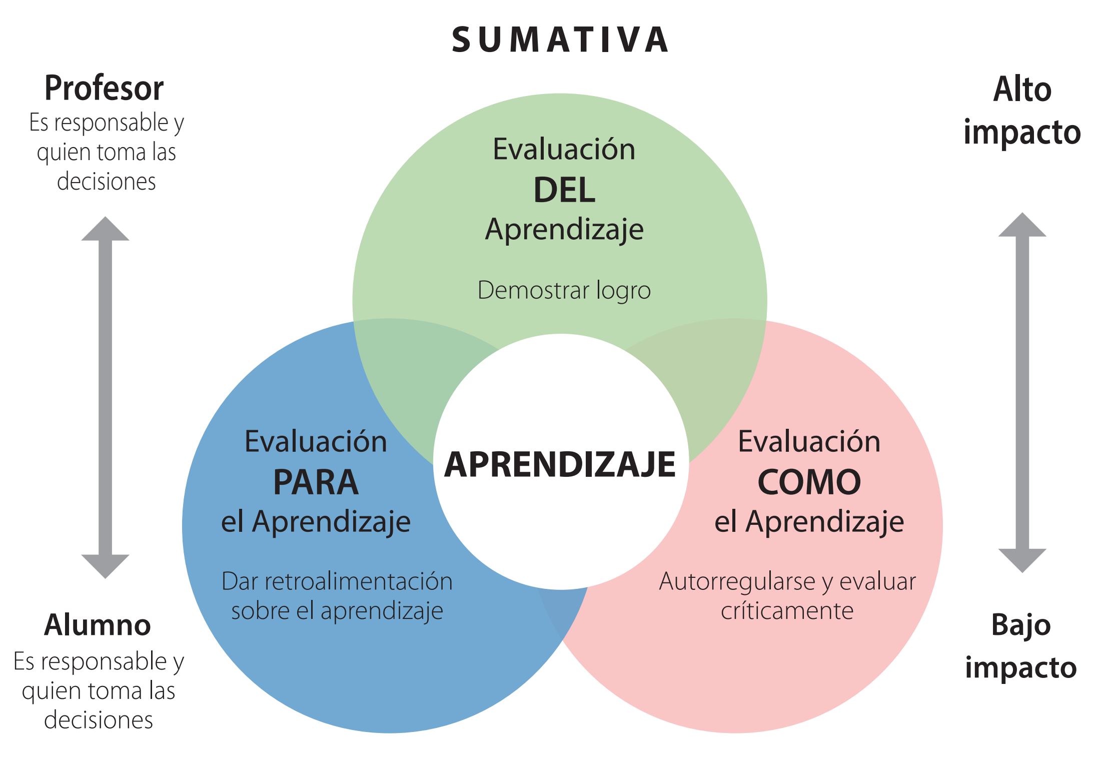


Tabla 2. Características de la evaluación del, para y como aprendizaje (adaptado de Harapnuik, 2021)

Evaluación	Del aprendizaje	Para el aprendizaje	Como aprendizaje
Tipo	Sumativa	Formativa	Formativa
Qué	El profesorado determina el progreso del estudiante o la aplicación de conocimientos y habilidades comparados con un estándar.	Docentes y pares valoran el progreso y el aprendizaje para ayudar a los estudiantes a mejorar.	El educando toma responsabilidad de su propio aprendizaje y hace preguntas sobre el proceso, explora cómo mejorar.
Quién	Profesor(a).	Profesor(a) y pares	Estudiante y pares.
Cómo	Evaluaciones formales para colectar evidencia del progreso del estudiante, se usa para medir logro y otorgar calificaciones.	Actividades de evaluación formales e informales como parte del aprendizaje, para mejorarlo e informar la planeación del aprendizaje futuro.	Los estudiantes usan la retroalimentación formal e informal, así como la auto-evaluación, para ayudar a su comprensión de los siguientes pasos en su aprendizaje. Usan además los exámenes como herramienta de aprendizaje.
Cuándo	Exámenes y reportes periódicos.	Retroalimentación continua	Práctica reflexiva continua
Por qué	Clasificar, documentar, verificar, promover.	Mejorar el aprendizaje.	Lograr aprendizaje más profundo, aprender a aprender.
Énfasis	Puntuaciones, calificaciones, competitividad.	Retroalimentación, apoyo, colaboración.	Colaboración, reflexión, auto-evaluación.

Después de analizar por separado estos tres elementos de la evaluación, es aparente que forman parte de un modelo conceptual entrelazado con interacciones dinámicas y cierto grado de superposición entre ellos, en el que las tres categorías (del, para y como) no son mutuamente excluyentes. La Figura 2 ejemplifica sus relaciones y puntos de contacto, así como el rol cambiante de docentes, estudiantes e impacto de las evaluaciones en cada uno de ellos. Por ejemplo, las evaluaciones de bajo impacto, como la que hacemos a las participaciones en clase de los estudiantes, cuyos propósitos son diagnósticos, motivacionales y para retroalimentación, son un ejemplo de superposición entre los círculos de evaluación del y para el aprendizaje. Cuando los estudiantes reciben tareas o exámenes para auto-eva-

luación, hay superposición entre evaluación del y como aprendizaje. Existen varias combinaciones potenciales entre estas tres grandes estrategias de evaluación y sus implicaciones para el aprendizaje, pero el concepto central que las une es que todas ellas deben facilitar el aprendizaje de los estudiantes.

Cada componente del modelo tiene sus particularidades, pros y contras, énfasis en diferentes tiempos y propósitos, pero cada uno tiene un rol clave en el aprendizaje de los estudiantes y el éxito del proceso educativo. La comprensión de estos elementos y sus interacciones por el profesorado, el estudiantado y las autoridades educativas, permitirá un diseño e implementación más adecuada de las experiencias de aprendizaje. Eventualmente sería deseable que la necesidad del uso diferenciado de estos términos pase a la historia, ya que idealmente toda la evaluación debería ser para el aprendizaje. En palabras de Black y Wiliam (1998), "el usuario último de la información de evaluación que se obtiene para mejorar el aprendizaje es el estudiante", por ello es fundamental que los estudiantes no sean ajenos a las actividades de evaluación que diseñan e implementan las universidades y el profesorado. La evaluación debe motivar al estudiantado y apoyarle en su transición a ser un aprendiz para toda la vida.

EVALUACIÓN AUTÉNTICA

Uno de los principales retos de la evaluación, sobre todo en escenarios artificiales estandarizados controlados, es la poca autenticidad de los exámenes escritos de baja fidelidad, por lo que es importante comentar el concepto de la evaluación auténtica. El término "evaluación auténtica" no es un tipo separado de evaluación, sino más bien un adjetivo que denota características de la misma que se acercan a los retos que las personas enfrentan en la vida cotidiana, en el trabajo y el ejercicio profesional, y que promueve el aprendizaje (McKean y Aitken, 2016). Este tipo de evaluación pretende replicar las tareas y el desempeño encontrados típicamente en el mundo del trabajo, y se ha documentado que tiene un impacto positivo en el aprendizaje, la autonomía, la motivación, la meta-cognición y la auto-regulación (Villarroel et al, 2018). Autenticidad se entiende como realismo, que implica enlazar el conocimiento con la vida diaria y el trabajo, en contextos relevantes, con problematización de los retos a resolver definidos en el currículo. Este tipo de evaluación intenta integrar lo que ocurre en la escuela y el aula con el empleo y el escenario del trabajo.

Se han descrito criterios para que una evaluación sea auténtica: representatividad (que evalúe adecuadamente el panorama del dominio); significado (tareas relevantes y pertinentes a la vida real); complejidad cognitiva (habilidades de orden superior); y cobertura de contenido (material incluido amplio) (Ashford-Rowe et al, 2014; McKean y Aitken, 2016). En una revisión sistemática del tema, Villarroel y colaboradores identificaron 13 características consistentes en la evaluación auténtica, que se pueden organizar en tres dimensiones conceptuales (Tabla 3).

Tabla 3. Dimensiones y características de la evaluación auténtica (adaptado de Villarroel et al, 2018)

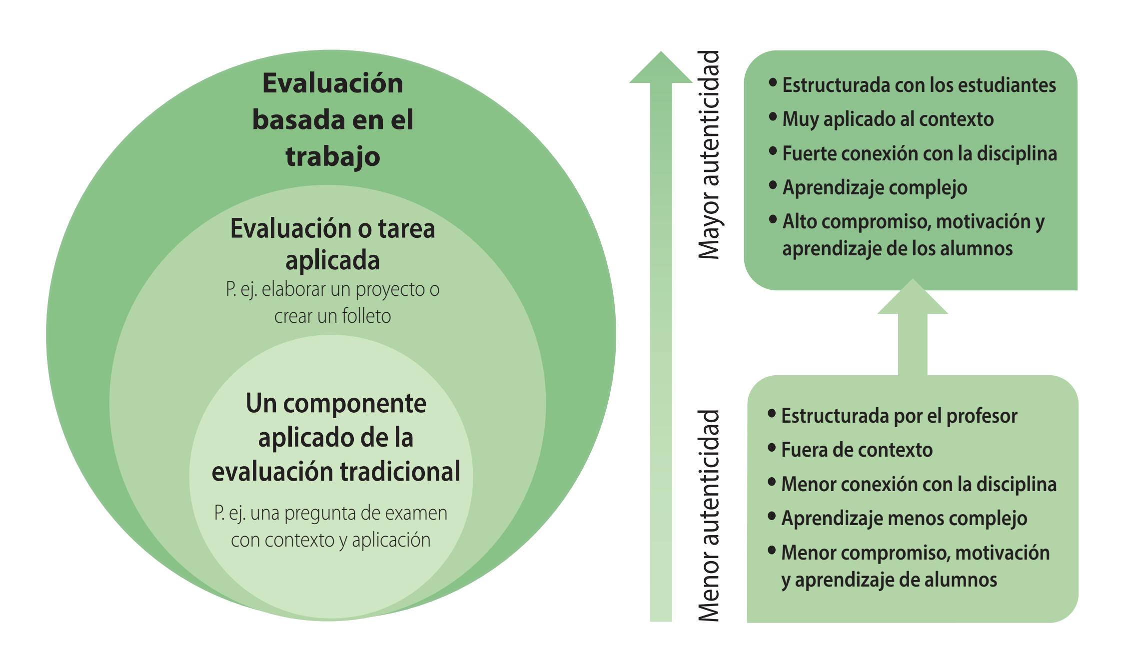
Dimensiones	Características
Realismo	• Problemas contextualizados de la vida diaria. • Relevancia más allá del salón de clase. • Desempeño auténtico. • Competencias para el desempeño en el trabajo. • Tareas similares al mundo real. • Valor práctico.
Reto cognitivo	• Pensamientos de orden superior. • Habilidad para resolver problemas. • Habilidad para tomar decisiones.
Juicio evaluativo	• Retroalimentación. • Sentimiento formativo. • Conocimiento a priori de los criterios de evaluación.

La evaluación auténtica ha adquirido cada vez más relevancia en el mundo moderno, en el que los empleadores identifican carencias y limitaciones en las habilidades para resolver problemas de los graduados de las universidades. Esto ha motivado que se incrementen los métodos de evaluación que utilizan esta perspectiva, combinando el uso de la tecnología, agregando escenarios y viñetas que den contexto y problematización a los reactivos de los exámenes escritos, utilizando centros de simulación con simuladores reales o virtuales, entre otras herramientas e instrumentos.

A pesar de sus beneficios, existen barreras para la adopción generalizada de la evaluación auténtica que deben tomarse en cuenta: la tradición en algunas universidades en las que los exámenes escritos de conocimiento decontextualizados son la norma; la dificultad para incorporar los principios de evaluación auténtica en el día a día de la docencia y evaluación, por cuestiones de tiempo y recursos; así como la limitada formación docente sobre el tema. De cualquier forma, es fundamental que en las evaluaciones modernas se haga el esfuerzo de incorporar la evaluación auténtica, para mejorar el aprendizaje de los estudiantes y su aceptación de las evaluaciones. Al sentir que los exámenes son más realistas y apegados a los

problemas que enfrentarán en la vida real, estudiantes y docentes se sentirán más satisfechos y motivados a participar en el proceso. Por otra parte, se debe conceptualizar a la autenticidad como un continuo, con niveles variables de fidelidad (Figura 3). Este esquema ayuda a pensar en el diseño de evaluaciones que pueden ir de un nivel menor a uno mayor de autenticidad y realismo, de cualquier forma, diseñar pruebas y exámenes requiere que estos estén alineados con las metas de aprendizaje del currículo y los métodos de enseñanza de la asignatura o curso. La evaluación auténtica es fuertemente determinada por la disciplina o área del conocimiento de la profesión que se estudia (medicina, música, leyes, enfermería, etc.), ya que el contexto de la práctica influye en la percepción de la importancia y legitimidad de lo que se debe aprender. Los estudiantes anticipan los escenarios en donde aplicarán lo aprendido en la universidad.

Figura 3. El continuo de la evaluación auténtica (adaptado de NFETLHE, 2017).



EVALUACIÓN ALTERNATIVA

El término de evaluación alternativa se ha usado con mayor frecuencia en los últimos años, y es particularmente importante en el contexto actual de la pandemia por COVID-19, en el que ha habido un movimiento amplio y profundo hacia la evaluación en línea, a distancia y mediada por tecnología (Jiao y Lissitz, 2020). El tema de la evaluación en línea y sus implicaciones en la pandemia es analizado en el capítulo 9. En este apartado planteamos el concepto de evaluación alternativa y sus implicaciones en el proceso educativo. Este término se refiere a una aproximación a la evaluación en forma diferente a la que se ha hecho tradicionalmente

(generalmente exámenes escritos con papel y lápiz). Stobart y Gipps describen tres niveles de evaluación alternativa que se refieren a las diferentes facetas del concepto: en primera instancia, en el nivel más básico, puede referirse a formatos alternativos, por ejemplo, el uso de una computadora en lugar del examen con papel y lápiz, sin modificar el contenido de la prueba, o usar preguntas abiertas en lugar de reactivos de opción múltiple; el segundo nivel puede ser el uso de un enfoque diferente a la evaluación, por ejemplo evaluación del desempeño, uso de portafolios; y el tercero sería el uso de evaluación formativa para el aprendizaje, en lugar (o complementario a) la evaluación sumativa del aprendizaje con exámenes de alto impacto, como se ha comentado en este capítulo.

Más allá de utilizar el término de evaluación alternativa como algo diferente a lo que se hace tradicionalmente, es importante revisar nuestras premisas de lo que es la educación, enseñanza, aprendizaje y evaluación, y explorar evaluaciones alternativas que pueden tener diferentes premisas. Anderson (1998) propuso que la evaluación alternativa tiene los siguientes supuestos: existen múltiples significados de lo que es el conocimiento; el aprendizaje es construido socialmente y activamente; tanto el proceso como el producto son relevantes; debe enfocarse en habilidades multimodales; incluye la subjetividad; el control del aprendizaje es compartido por profesor y estudiante, como un proceso colaborativo; y primordialmente, el propósito primario es facilitar el aprendizaje.

CONCLUSIONES

El término evaluación en educación incluye un abanico amplísimo de conceptos, estrategias e instrumentos, cuya comprensión y aplicación tiene enorme potencial para mejorar el proceso educativo. Es necesario que docentes, estudiantes y autoridades se involucren en asimilar los principios básicos de esta temática, para generar un andamiaje de apoyo que facilite el aprendizaje de los estudiantes, lo que implica apropiarse de los instrumentos, conceptos y estrategias de la evaluación del, para y como aprendizaje. Citando a un académico mexicano, el Dr. Tiburcio Moreno, la evaluación tiene muchas caras, y en países como el nuestro ha estado permeada por una visión empirista que descansa en el principio: "*Todos sabemos de evaluación, porque alguna vez hemos sido evaluados*" (Moreno Olivos, 2010). Debemos mejorar nuestros conocimientos y habilidades en evaluación, es una obligación ética y moral de todo el profesorado.

REFERENCIAS

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, y Joint Committee on Standards for Educational and Psychological Testing. (2014). *Standards for educational and psychological testing*. Washington, D.C.: AERA. https://www.testingstandards.net/open-access-files.html

Anderson, R. S. (1998). Why talk about different ways to grade? The shift from traditional assessment to alternative assessment. *New Directions for Teaching and Learning*, 74, 5–16.

Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, *39*(2), 205-222. https://doi.org/10.1080/02602938.2013.819566

Au, W. (2007). High-Stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, *36*(5), 258-267.

Bennett, R. E. (2015). The Changing Nature of Educational Assessment. *Review of Research in Education*, *39*(1), 370–407. https://doi.org/10.3102/0091732X14554179

Black, P., Wiliam, D. (1998) Assessment and Classroom Learning, *Assessment in Education: Principles, Policy & Practice*, 5:1, 7-74, DOI: 10.1080/0969595980050102

Boursicot K., Etheridge L., Setna Z., Sturrock A., Ker J., Smee S., et al. (2011). Performance in assessment: consensus statement and recommendations from the Ottawa conference. *Med Teach*, *33*(5), 370-383. DOI: 0.3109/0142159X.2011.565831

Brennan, R. L. (2006). Perspective on the Evolution and Future of Educational Measurement. En R. L. Brennan (Ed.), *Educational Measurement. National Council on Measurement in Education and American Council on Education* (pp. 1-16). Westport, Connecticut: Praeger Publishers.

Carter, K. (1984). Do teachers understand principles for writing tests? *Journal of Teacher Education*, *35*(6), 57-60.

Freud, S. (1937). Analysis Terminable and Interminable. *Int. J. Psycho-Anal.*, 18:373- 405.

Green, M., Moeller, J. y Spak, J. (2018). Test-enhanced learning in health professions education: A systematic review: BEME Guide No. 48. *Medical Teacher, 40(*4), 337-350. https://doi.org/10.1080/0142159X.2018.1430354

Harapnuik, D. K. (2021). Assessment OF/FOR/AS Learning. En blog: It's About Learning. Creating Significant Learning Environments. https://www.harapnuik.org/?page_id=8900

Hodges B. (2013). Assessment in the post-psychometric era: learning to love the subjective and collective. *Medical teacher*, *35*(7), 564–568. https://doi.org/10.3109/0142159X.2013.789134

Jiao, H. and Lissitz, R.W. (2020), What Hath the Coronavirus Brought to Assessment? Unprecedented Challenges in Educational Assessment in 2020 and Years to Come. Educational Measurement: Issues and Practice, 39: 45-48. https://doi.org/10.1111/emip.12363

Lok, B., McNaught, C., & Young, K. (2016) Criterion-referenced and norm-referenced assessments: compatibility and complementarity, *Assessment & Evaluation in Higher Education*, *41*(3), 450-465, DOI: 10.1080/02602938.2015.1022136

Maki, P. L. (2010). *Assessing for Learning: Building a Sustainable Commitment across the Institution*. Stylus Publishing, LLC.

Man Sze Lau, A. (2016). "Formative good, summative bad?" – A review of the dichotomy in assessment literature. *Journal of Further and Higher Education*, *40*(4), 509-525. DOI: 10.1080/0309877X.2014.984600

Martínez Rizo, F. (2009). Evaluación formativa en aula y evaluación a gran escala: hacia un sistema más equilibrado. *Revista Electrónica de Investigación Educativa*, *11*(2). Recuperado de http://redie.uabc.mx/redie/article/view/231

Martone, A., Sireci, S. G. (2009). Evaluating Alignment Between Curriculum, Assessment, and Instruction. *Review of Educational Research*, *79*(4), 1332–1361.

McKean M., Aitken E.N. (2016) Educational Renovations: Nailing Down Terminology in Assessment. In: Scott S., Scott D., Webber C. (eds) *Leadership of Assessment, Inclusion, and Learning. The Enabling Power of Assessment*, *3*. Springer, Cham. https://doi.org/10.1007/978-3-319-23347-5_2

Miller, M. D., Linn, R. L. y Gronlund, N. E. (2012). *Measurement and Assessment in Teaching* (11va ed.). Londres: Pearson.

Montgomery, P.C., & Connolly, B.H. (1987). Norm-referenced and criterion-referenced tests. Use in pediatrics and application to task analysis of motor skill. *Physical Therapy, 67 12*, 1873-6.

Moreno-Olivos, T. (2010). Lo bueno, lo malo y lo feo: las muchas caras de la evaluación. *Revista Iberoamericana de Educación Superior, I*(2), 84-97. https://www.ries.universia.unam.mx/index.php/ries/article/view/6

Mueller, J. (2016). Authentic Assessment Toolbox. North Central College, Naperville, IL. http://jfmueller.faculty.noctrl.edu/toolbox/index.htm

National Forum for the Enhancement of Teaching and Learning in Higher Education (2017a). Assessment OF/FOR/AS Learning. https://www.teachingandlearning.ie/our-priorities/student-success/assessment-of-for-as-learning/

National Forum for the Enhancement of Teaching and Learning in Higher Education (2017b). Authentic Assessment in Irish Higher Education. https://hub.teachingandlearning.ie/wp-content/uploads/2021/06/NF-2017-Authentic-Assessment-in-Irish-Higher-Education.pdf

Newble, D. I., y Jaeger, K. (1983). The effect of assessments and examinations on the learning of med dical students. *Med Educ*., *17*(3), 165-71.

Norcini, J., Anderson, B., Bollela, V., Burch, V., Costa, M. J., Duvivier, R., (2011). Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med. Teach*., *33*(3), 206-214.

Roediger, H. L. III, Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. P. Mestre & B. H. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (pp. 1–36). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00001-6

Rowntree, D. (1977). *Assessing students: How shall we know them?* Londres: Kogan Page Ltd.

Sánchez-Mendiola, M., Delgado-Maldonado, L. (2017). Exámenes de alto impacto: Implicaciones educativas. *Inv Ed Med., 6*(21), 52-62. DOI: 10.1016/j.riem.2016.12.001

Stobart, G., & Gipps, C. (2010). Alternative assessment. In B. McGraw, E. Baker, & P. Peterson (Eds.), *International Encyclopedia of Education* (3rd ed., pp. 202–208). Oxford, UK: Elsevier.

UNESCO (2020). COVID-19. A glance of national coping strategies on high-stakes examinations and assessments. Paris: UNESCO. https://en.unesco.org/sites/default/files/unesco_review_of_high-stakes_exams_and_assessments_during_covid-19_en.pdf

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018) Authentic assessment: creating a blueprint for course design, Assessment & Evaluation in Higher Education, *43*(5),840-854, DOI:10.1080/02602938.2017.1412396

Wiliam, D. (2011). What is assessment for learning?, *Studies in Educational Evaluation*, 37(1), 3-14, https://doi.org/10.1016/j.stueduc.2011.03.001.

Capítulo 2

VALIDEZ, CONFIABILIDAD Y AMENAZAS A LA VALIDEZ

Blanca Ariadna Carrillo Ávalos, Melchor Sánchez Mendiola

"La validez es simple. La validación puede ser difícil."
Michael Kane, 2009

INTRODUCCIÓN

A lo largo de la vida como docentes realizamos muchas evaluaciones para intentar conocer el nivel de conocimiento o desempeño de nuestros estudiantes. Este proceso implica la elaboración, aplicación e interpretación de exámenes de diferentes tipos: diagnósticos, formativos y sumativos. Independientemente de su finalidad, la meta de cualquier evaluación incluye la identificación del nivel de algún constructo, como conocimiento sobre química orgánica, habilidad para resolver problemas de trigonometría, o competencia de comunicación escrita. Los resultados de las evaluaciones idealmente deben reflejar de una manera precisa y reproducible lo que se pretende evaluar, para poder interpretar de forma racional los resultados de la misma y estar en capacidad de realizar inferencias y tomar decisiones con fundamentos sólidos. Cuando evaluamos a nuestros estudiantes sobre un tema particular, deseamos identificar el proceso y resultados del aprendizaje que permitan inferir el nivel de desempeño en los constructos de interés. Después de aplicar las evaluaciones, obtenemos resultados en forma de puntuaciones que ayudan a tomar decisiones, que conllevan las siguientes interrogantes: ¿estamos evaluando exactamente lo que deseamos evaluar?, ¿qué implican los resultados con respecto al avance académico del alumno?, si se trata de una evaluación sumativa, ¿cuál es la calificación mínima para aprobar el curso?, ¿qué tan reproducible es la medición?, entre muchas otras.

La evaluación en educación es una disciplina cada vez más sofisticada y sustentada en investigación, que requiere incorporar conceptos académicos fundamentales para llevarse a cabo con profesionalismo y solidez metodológica (Instituto Nacional para la Evaluación de la Educación, 2017). El pilar conceptual más importante de la evaluación en educación es la validez, tema del que trata este capítulo. Actualmente el concepto de **validez** ha evolucionado del tradicional "medir lo que se pretende medir", a un modelo más amplio y profundo, en

el que *"se refiere al grado en que la evidencia y la teoría respaldan las interpretaciones de los puntajes de una prueba para usos propuestos de las pruebas"* (AERA, APA y NCME, 2018). Es un conjunto de acciones que se ubican a lo largo del proceso evaluativo, para fundamentar la interpretación de los resultados y así generar inferencias. El análisis de validez, o **validación**, es el proceso mediante el que evaluamos las evidencias que se presentan para determinar cuál es el grado de validez (Cook y Hatala, 2016). Se puede realizar para los diferentes tipos de exámenes, diagnósticos, formativos y sumativos, aunque es particularmente relevante para las evaluaciones sumativas de alto impacto.

Tradicionalmente la validez en educación se clasificaba como "las 3 Cs": validez de contenido, de criterio y de constructo (Cronbach y Meehl, 1955). En la definición actual esta distinción desapareció, ya que el modelo vigente propone diversas fuentes de evidencia que arrojan luz sobre distintos aspectos de la validez, no es que reflejen diferentes tipos de la misma. La validez es un concepto unitario, por lo que se considera que toda la validez es validez de constructo. La palabra **constructo** significa colecciones de conceptos abstractos y principios, inferidos de la conducta y explicados por una teoría educativa o psicológica, es decir, atributos o características que no pueden observarse directamente, por ejemplo: inteligencia emocional, extroversión, conocimientos sobre matemáticas (AERA, APA y NCME, 2018).

La validez es un juicio valorativo holístico e integrador que requiere múltiples fuentes de evidencia para su interpretación, e intenta responder a la pregunta: *"¿qué inferencias pueden hacerse sobre la persona basándose en los resultados del examen?"* (Downing, 2003). Validez no es una propiedad intrínseca del examen, sino del significado de los resultados en un entorno educativo específico y las inferencias que pueden hacerse a partir de los mismos (Carrillo-Ávalos et al., 2020; Downing, 2003). Por ejemplo, los resultados de los exámenes de admisión a las universidades no deben interpretarse categóricamente como evidencia predictiva de que la persona vaya a tener éxito en la vida, ya que el examen no está diseñado con ese propósito. Quienes desarrollan pruebas e interpretan sus resultados, deben poseer nociones básicas de los conceptos de validez y confiabilidad, para incorporarlos de manera apropiada en el proceso de enseñanza y aprendizaje. Este capítulo presenta un panorama básico de dichos conceptos, además de describir las principales amenazas a la validez.

VALIDEZ

Anteriormente se hablaba de tres tipos de validez: de contenido, de constructo y de criterio; la de criterio se dividía en validez concurrente y validez predictiva (Cronbach y Meehl, 1955). Posteriormente, a finales del siglo XX, un nuevo marco de referencia de validez fue propuesto y aceptado por las principales organizaciones de evaluación educativa y pruebas psicológicas (American Educational Research Association et al., 2018), incorporando el concepto holístico de validez de constructo. Este modelo establece que, para determinar el grado de validez de los usos e interpretaciones de los resultados de una evaluación, se deben

proveer diversos elementos que lo demuestren (Downing, 2003). Este esquema propone los siguientes elementos como **cinco fuentes de evidencia** de validez (Downing, 2003; Messick, 1989):

1) **Evidencia basada en el contenido de la prueba.** El contenido de la prueba alude a los conocimientos que evalúa; por ejemplo, en el caso de un examen final de un curso de biología evolutiva, son los relacionados con la historia de la teoría de la evolución y sus evidencias, sus bases genéticas, los procesos del cambio evolutivo, el origen de las especies, macro-evolución y la evolución de los homínidos. Es decir, todos los temas que debe saber el alumnado cuando termina de estudiar esta asignatura.

   Para demostrar que los usos e interpretaciones de un examen de esta naturaleza son adecuados, podemos buscar evidencia de que, en efecto, las preguntas del examen versan sobre los temas mencionados. Estas evidencias descansan en cuatro elementos (Sireci y Faulkner-Bond, 2014):

   - *Definición del dominio.* Se trata de la descripción detallada de las áreas del contenido y las habilidades cognitivas que se desean evaluar del constructo definido en el currículo, o de los resultados de la actividad de aprendizaje. En la tabla de especificaciones de la prueba se deben enlistar las subáreas del contenido y los niveles cognitivos que estaremos midiendo (conocimiento, comprensión, aplicación, análisis, síntesis o evaluación), así como los estándares del contenido y los objetivos curriculares.
   - *Representación del dominio*. Con frecuencia se hacen demasiadas preguntas sobre un tema y se dejan de lado otros. Para decidir cuántos ítems, preguntas o reactivos corresponden a cada tema, podemos establecer una tabla de ponderaciones en la que se establece la relevancia de cada uno, a los de mayor importancia se asignan más preguntas en el examen. También se debe buscar la alineación entre el contenido del examen y los estándares para ese conocimiento. Por ejemplo, cada asignatura que cursa un alumno tiene una carta descriptiva en la que se han establecido los objetivos o metas de aprendizaje, lo que se debe saber al terminar de cursar esa materia. En un examen sumativo, las preguntas deberán contemplar tales objetivos y coincidir con la tabla de especificaciones. Otra manera de demostrar alineación es que expertos en el tema opinen sobre el examen, cómo los ítems evalúan los diferentes aspectos del constructo de interés.
   - *Relevancia del dominio*. Se refiere a qué tan importantes son los ítems con respecto al aspecto del constructo que se está midiendo, que se pregunten conceptos importantes y no datos triviales.
   - *Procedimientos apropiados de diseño de la prueba*. Los procedimientos que se llevan a cabo al diseñar la prueba deben servir para asegurarse de que su contenido evalúa fielmente y representa por completo al constructo de interés. Esto puede comprobarse al implementar controles de calidad durante el desarrollo del examen, como revisión de ítems por expertos en contenido para asegurarse de su veracidad técnica,

revisión por expertos en evaluación para verificar que estén bien elaborados, probar los ítems por medio de estudios piloto.

- *Credenciales de los creadores del examen, elaboradores de reactivos y expertos en contenido*. Es importante documentar que las personas que intervienen en el proceso de diseño del examen, definición del constructo, elaboración de los reactivos y análisis de los resultados, tengan las credenciales correspondientes para dar certidumbre a todo el proceso. Si el examen evalúa habilidades de traducción del idioma alemán, deben participar expertos en contenido que dominen dicho constructo; si se trata de elaborar un examen práctico de habilidades de comunicación verbal, deben participar expertos en el tema y profesionales del diseño de este tipo de exámenes.

2) **Evidencia basada en los procesos de respuesta.** Los procesos de respuesta son los procesos mentales que lleva a cabo el sustentante cuando contesta las preguntas de una prueba. Se esperaría que responda cada ítem integrando los conocimientos que se indagan y que no esté tratando de adivinar la respuesta correcta. Si el examen contiene preguntas de opción múltiple (POM) o abiertas, cuando el alumno termine de contestar se puede indagar cómo llegó a la respuesta a través de una entrevista cognitiva, así se puede saber si comprendió los conceptos clave, si hay errores de redacción en las preguntas, cuál fue el razonamiento que utilizó, y si hay posibilidad de que existan falsos positivos (que el alumno haya utilizado un razonamiento erróneo para llegar a la respuesta). Al final, lo que se busca es que el estudiante aplique en verdad los conocimientos adquiridos para resolver los problemas que se proponen en el examen y que luego los pueda aplicar en la vida real. Otra manera de aportar a esta evidencia es a través de modelos matemáticos que evalúen la dificultad y el tiempo que tardan en contestar cada ítem (Padilla y Benítez, 2014).

   Algunos autores agregan en este apartado la familiaridad de los sustentantes con el formato del examen (por ejemplo, evaluación asistida por computadora), la validación de la hoja de respuestas correctas, el control de calidad del reporte de los resultados, entre otros (Downing, 2003).
3) **Evidencia basada en la estructura interna.** La estructura interna presenta tres características básicas: dimensionalidad, funcionamiento diferencial y confiabilidad (Rios y Wells, 2014). Al diseñar la prueba, se debe determinar cuáles dimensiones se desean evaluar sobre el constructo de interés, y esta información se describe en la tabla de especificaciones del examen. Siguiendo el ejemplo de un examen de biología evolutiva, una dimensión sería el proceso cognitivo que siguen los alumnos para resolver problemas de ese tema. Sin embargo, en ocasiones los exámenes contienen ítems que evalúan diferentes dimensiones del mismo constructo, por ejemplo, procesos cognitivos y valores. Para saber cuántas dimensiones posee el examen, se pueden hacer diversos análisis, entre ellos un análisis factorial confirmatorio, el que también permite identificar cuáles ítems valoran la misma dimensión buscando la relación existente entre ellos. Con el resultado de este análisis podemos justificar, por ejemplo, dar el mismo valor a todos los ítems, porque todos evalúan la misma dimensión del mismo constructo.

Las pruebas también deben aportar resultados imparciales, por lo que es útil buscar información con respecto al funcionamiento diferencial de los ítems (DIF, por sus siglas en inglés) (Leenen, 2014; Rios y Wells, 2014). Se trata de que las características de los ítems de la prueba son comparables entre diferentes grupos; es decir, que no habrá diferencia entre los resultados de hombres y mujeres, o por edades, por ejemplo. Para conocer el grado de invarianza de los resultados entre grupos, se pueden realizar pruebas basadas en el funcionamiento diferencial del ítem, como un análisis factorial confirmatorio de grupos múltiples.

Este apartado de evidencia de validez también incluye el análisis estadístico de los reactivos de la prueba (Downing, 2003), para documentar diversos indicadores como grado de dificultad, índice de discriminación, confiabilidad, error estándar de medición, curvas características del ítem, entre muchos otros, para lo cual existen diversos modelos y métodos matemáticos que se describen en otros capítulos de esta obra.

4) **Evidencia basada en las relaciones con otras variables.** En el ejemplo mencionado, los alumnos que hicieron el examen de la unidad de las bases genéticas de la evolución en la asignatura de biología evolutiva, también contestaron uno de genética, en donde se les hicieron preguntas acerca de genética de poblaciones. Ambas pruebas son semejantes en cuanto al constructo evaluado. La validación de la interpretación de los resultados del examen de biología evolutiva podría incluir el análisis de la relación convergente con los resultados del examen de genética, si ambas pruebas valoran constructos similares. Por otro lado, también puede existir una relación divergente cuando se evalúan constructos diferentes. El análisis para establecer ambas correlaciones se puede realizar por medio de la matriz multirasgo-multimétodo (MTMM) de Campbell (Campbell y Fiske, 1959), en el que se establece una matriz de correlaciones entre las dos pruebas en donde se representan al menos dos características y cada una de ellas debe ser medida por lo menos por dos métodos.

Esta fuente de evidencia proporciona información acerca del grado en que la relación divergente o convergente es coherente con el constructo cuya medición es la base de la interpretación de los resultados de la prueba. Otra evidencia que aporta a las relaciones con otras variables es la relación entre la prueba y el criterio, la que se puede establecer por medio de uno de estos diseños:

- *Estudio predictivo*. Para conocer el grado de relación entre el resultado de la prueba y el resultado del criterio que se evalúa posteriormente. Por ejemplo, si en el examen de admisión a la universidad se evalúa biología general, podríamos tratar de responder si las calificaciones de biología general en el examen de admisión predicen las de biología evolutiva. En este caso, el criterio sería el resultado de biología evolutiva. Por otro lado, también se pueden hacer estudios para hacer predicciones diferenciales por grupo de edad, sexo, antecedentes académicos, entre otras.
- *Estudio concurrente*. Se mencionó en el ejemplo que los alumnos llevan las asignaturas de genética y biología evolutiva al mismo tiempo, por lo que las evaluaciones

correspondientes también ocurrieron en fechas cercanas. Si se evalúa un constructo al mismo tiempo, podemos estimar la relación entre las puntuaciones de la prueba y del criterio.

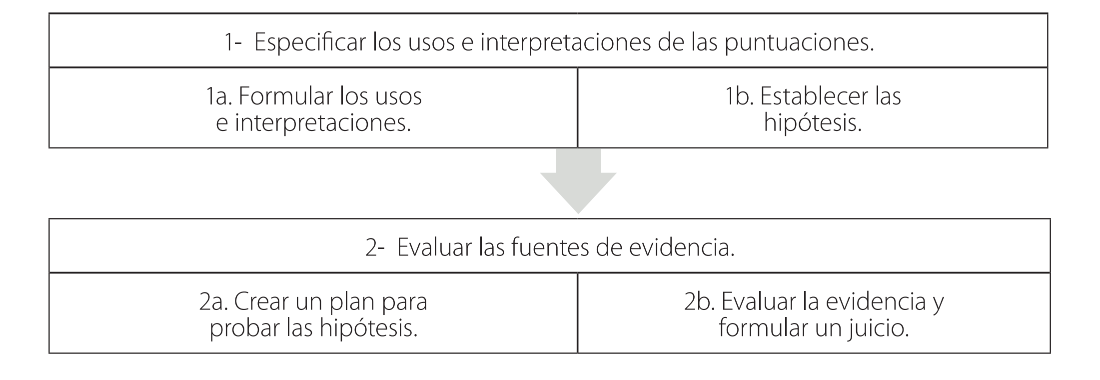
5) **Evidencia basada en las consecuencias de la prueba.** Los resultados de las pruebas, sobre todo las sumativas y de alto impacto como los exámenes de admisión o de titulación, tienen grandes consecuencias para los sustentantes. En el caso de los exámenes de admisión a las universidades, uno de los objetivos puede ser seleccionar a los sustentantes con mayor probabilidad de tener éxito académico (Downing, 2003). Las consecuencias de pruebas sumativas de impacto moderado son importantes también, ya que se van agregando para generar un promedio final que permita el avance en las trayectorias escolares.

   Las fuentes de evidencia de consecuencias de las pruebas se buscan al analizar cuáles son los impactos de los resultados en el avance académico del estudiante, en sus familias, y en la sociedad. En las pruebas cuyos resultados se reportan con base en criterio, la decisión del punto de corte mínimo debe fundamentarse; en el caso de pruebas con referencia a norma también se debe sustentar la decisión de, por ejemplo, por qué admitir a los 100 sustentantes con las puntuaciones más altas. Otras fuentes de evidencia en este rubro pueden ser las consecuencias de aprobar o reprobar, de falsos positivos y falsos negativos, y las consecuencias institucionales. La manera de obtener esta información puede ser a través de entrevistas, grupos focales, cuestionarios, para conocer cuáles son los componentes más importantes de los programas académicos y sus puntos de mayor impacto (Lane, 2014).

VALIDACIÓN

La validación es un proceso que se debe planear al mismo tiempo que se diseña la prueba, para asegurarse de contar con las fuentes de evidencia necesarias para obtener el mayor grado posible de validez de la interpretación de sus resultados. Una manera de realizar este proceso se sugiere a continuación (Figura 1).

Figura 1. Esquema de pasos generales para el proceso de validación de una prueba (elaboración propia)



1. Especificar los usos e interpretaciones de las puntuaciones

1a. Formular los usos e interpretaciones. Los usos y las interpretaciones de las puntuaciones que se obtienen en una prueba son conceptos diferentes y ambos se deben aclarar desde que inicia el diseño de la prueba. La justificación del uso de las puntuaciones se puede conocer respondiendo a preguntas como: ¿debemos utilizar estas puntuaciones para tomar decisiones sobre quiénes pueden ingresar a un programa de posgrado? Para ello, se deben conocer las características de los usuarios principales, quiénes son las personas que presentan dichas evaluaciones; además, también son de interés las instituciones que las desarrollan, profesores, personal administrativo.

En cuanto a las interpretaciones, se pueden determinar al responder preguntas como: ¿las calificaciones del examen de graduación reflejan las competencias que debe poseer un egresado de este programa de posgrado?, ¿los resultados de esta prueba miden el nivel de desempeño de los estudiantes en esta asignatura?

Los datos que se deben evaluar en este paso se describen en la Tabla 1.

Tabla 1. Formulación de los usos e interpretaciones de una prueba o examen

Dato necesario	Pregunta a contestar	Ejemplo
El objetivo de la prueba .	¿Para qué se aplica la prueba?	Establecer el nivel de desempeño en los conocimientos acerca de Biología Evolutiva.
Los usuarios propuestos .	¿Quiénes utilizarán los resultados de la prueba?	La Facultad de Ciencias Biológicas de la universidad.
Los usos.	¿Para qué se utilizarán las puntuaciones?	Las calificaciones serán utilizadas para determinar cuáles alumnos demuestran un desempeño satisfactorio en la asignatura.
El constructo medido.	¿Cuál es constructo que se evalúa con esta prueba?	Los conocimientos en Biología Evolutiva.
Las interpretaciones de los resultados.	¿Los resultados serán interpretados con referencia a norma o a criterio? ¿Los resultados de esta prueba miden el nivel de desempeño de los estudiantes en esta asignatura?	Los resultados del examen final de la asignatura se interpretan con referencia a criterio, por lo que quienes obtengan una puntuación mayor a 6.0 tendrán un desempeño satisfactorio.
La población examinada.	¿Quiénes contestarán el examen?	Los alumnos de 7° semestre de la licenciatura en Biología. Porcentaje de hombres y mujeres, rango de edades.

1b. Establecer las hipótesis. Las hipótesis son preguntas que nos podemos hacer acerca de la evaluación que se está elaborando. Deben probarse por medio de las fuentes de evidencia mencionadas. Algunos ejemplos de hipótesis se describen en la Tabla 2.

Tabla 2. Ejemplos de hipótesis que pueden buscar comprobarse durante el proceso de validación

Hipótesis	Fuente de evidencia
Los contenidos evaluados coinciden con los objetivos de aprendizaje. Las preguntas utilizadas conforman una muestra adecuada del posible universo de preguntas sobre los mismos temas. La ponderación de cada tema evaluado es correcta. El nivel de complejidad de los ítems utilizados es adecuado con respecto al nivel de estudios.	Evidencia basada en el contenido de la prueba.
Los procesos mentales que llevaron a cabo los estudiantes son los esperados.	Evidencia basada en los procesos de respuesta.
El examen es unidimensional (si eso era lo planeado). Si es multidimensional, la calificación de cada dimensión es consistente con el constructo evaluado. La confiabilidad de los resultados es adecuada. Si se utilizan las puntuaciones de diferentes partes de la prueba como sub-puntuaciones, se especifica y se justifica cómo se combinan dichas sub-puntuaciones.	Evidencia basada en la estructura interna.
Existe relación entre los resultados de otras evaluaciones (que también poseen un nivel de confiabilidad adecuado) y la evaluación sumativa que examinan el mismo tema. Se toman en cuenta las variables confusoras que pueden sesgar los resultados para grupos específicos. Se demuestra que el desempeño en la prueba predice el desempeño en el criterio evaluado.	Evidencia basada en las relaciones con otras variables.
Las consecuencias negativas involuntarias no son graves y son menores a las consecuencias positivas.	Evidencia basada en las consecuencias de la prueba.

2. Evaluar las fuentes de evidencia

2a. Crear un plan para probar las hipótesis. Con base en las hipótesis seleccionadas, se buscan las fuentes de evidencia y se reúne la información correspondiente.

2b. Evaluar la evidencia y formular un juicio. En este último paso se evalúan todas las evidencias en orden y se establece el grado de validez de la interpretación de las puntuaciones de la prueba evaluada. Este grado dependerá de la calidad de las evidencias presentadas y también de las evidencias más importantes, según la prueba.

AMENAZAS A LA VALIDEZ

Además de analizar las fuentes de evidencia de validez, se sugiere identificar elementos que puedan afectar el grado de validez de los resultados de la evaluación. Este paso es importante porque da fortaleza a las decisiones que se toman con base en los resultados de la prueba. Los elementos que reducen el grado de validez se les denomina amenazas a la validez; se les llama así porque interfieren con la correcta interpretación de las puntuaciones (Carrillo-Ávalos et al., 2020; Downing y Haladyna, 2004). Estas amenazas pueden estar presentes en cualquier tipo de evaluación. En general, se reconocen dos tipos de amenazas a la validez: subrepresentación del constructo (SC) y varianza irrelevante al constructo (VIC) (Downing, 2003; Messick, 1989). Para explicar estos conceptos, partiremos de la teoría clásica de los tests (TCT). De acuerdo con esta, la puntuación que se obtiene a partir de las respuestas de un examen (puntuación observada o total = X) está compuesta por la suma de la puntuación verdadera (V) más el error aleatorio (E_a) (de Champlain, 2010; Schuwirth y van der Vleuten, 2011):

$$X = V + E_a$$

De esta manera, los factores que tienen un efecto sistemático sobre la puntuación observada "X" se agregan a la puntuación verdadera "V". Estos factores incluyen tanto al constructo de interés, como a otros factores sistemáticos que se presentan en todas las versiones del examen, pero cuya medición no es el objetivo; por ejemplo, que un reactivo de opción múltiple no tenga una respuesta correcta entre las opciones, por lo que todos los alumnos la contestarán mal. Por otra parte, el error aleatorio (E_a) está conformado por todas las condiciones que aportan variabilidad a la puntuación verdadera "V" de manera no sistemática, ya que son diferentes en cada ocasión que se aplica el examen y para cada persona, como el cansancio, el desvelo y el estrés. En realidad, no conocemos el valor del error aleatorio ni el de la puntuación verdadera, aunque con diversos métodos estadísticos se pueden estimar (Haladyna y Downing, 2004).

Con base en lo anterior, la puntuación verdadera se puede dividir en dos partes: la puntuación del constructo de interés (θ), más la puntuación causada por factores sistemáticos (E_s). Por lo tanto, la fórmula quedaría así (Haladyna y Downing, 2004):

$$X = \theta + E_s + E_a$$

A partir de esta fórmula podemos definir los dos tipos de amenazas a la validez. Una de ellas se presenta cuando se mide θ a través de ítems que representan *de forma incompleta* el dominio del conocimiento que queremos evaluar: esto es la subrepresentación del constructo. Por otro lado, la varianza irrelevante al constructo se asocia con el error sistemático E_s, que es causado por la medición involuntaria de elementos que no son el objetivo de la medición, e interfieren con la medición del constructo de interés y, por lo tanto, con la validez de la interpretación de los resultados (Downing y Haladyna, 2004; Haladyna y Downing, 2004; Mes-

sick, 1989). La varianza indeseable de X debida a E_a también aporta elementos para conformar una amenaza a la validez; sin embargo, sus factores conllevan a una confiabilidad baja, de lo que hablaremos en la siguiente sección de este capítulo.

A continuación, se describen ejemplos de los tipos de amenazas a la validez:

- **Subrepresentación del constructo (SC).** En el ejemplo del examen sumativo de biología evolutiva mencionado, supongamos que consta de 10 preguntas, que constituyen una muestra del universo de todas las preguntas posibles sobre el mismo tema y que miden el mismo constructo. Se considera que 10 ítems son pocos para evaluar a cabalidad los temas involucrados, por lo que no son suficientes para representar adecuadamente al constructo de interés ni a las dimensiones que deseamos evaluar. Otros problemas que podríamos encontrar en esta muestra es que se estén evaluando contenidos triviales, contenidos diferentes a los estudiados, que solo pregunten aspectos de un solo tema o que la confiabilidad sea insuficiente. Es evidente que estos problemas están relacionados con la evidencia de validez de contenido. Si aplicáramos este examen de 10 reactivos, las amenazas a la validez pueden ser: número insuficiente de preguntas, sesgo, concordancia del dominio, baja confiabilidad (Moreno et al., 2004).
- **Varianza irrelevante al constructo.** Esta amenaza tiene origen en el error sistemático E_s causado por una variable irrelevante al constructo de interés. Por ejemplo:

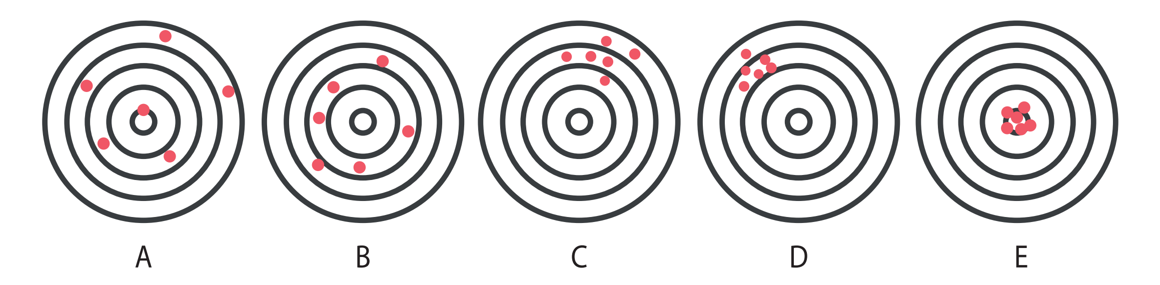
    - *Ítems defectuosos.* La escritura de un reactivo de opción múltiple de calidad debe cumplir con los estándares establecidos por expertos para evitar que aporten varianza irrelevante al constructo (Haladyna et al., 2002; National Board of Medical Examiners, 2016). La baja calidad de una pregunta la hace más difícil de contestar, convirtiéndola en un ítem defectuoso. Otros defectos posibles son que entre las opciones de respuesta se encuentre "todas las anteriores" o "ninguna de las anteriores", o el número de opciones (varios expertos consideran que el número óptimo es de tres a cuatro) (Abad et al., 2001; Haladyna et al., 2019; Rodríguez, 2005).
    - *Formato en negativo.* Varios expertos consideran que escribir preguntas en formato negativo debe evitarse porque se corre el riesgo de escribir un doble negativo (la pregunta y una o más de las opciones contienen términos negativos, por lo que identificar las opciones incorrectas implica negarlas, o sea, indicar que no son correctas) (Chiavaroli, 2017). También existe el riesgo de que el alumno no reconozca la negación en la pregunta, aunque se destaque en negritas "no" o "excepto", y que el proceso de respuesta llevado a cabo para responder no sea el esperado.
    - *Lenguaje.* Una pregunta que presenta una viñeta demasiado extensa o con explicaciones innecesarias ocasiona que el estudiante pierda tiempo leyéndola y en realidad puede evaluar la velocidad de lectura en lugar de medir el constructo deseado (Hicks, 2011; National Board of Medical Examiners, 2016). Por otro lado, las oraciones deben estar bien estructuradas, de manera que la pregunta quede clara.

- *Discordancia con el dominio*. Cuando no hay relación entre lo que se ha estudiado y lo que se está evaluando.
- *Ítems muy fáciles o difíciles*. Reactivos que seguramente todos los alumnos contestarán correctamente, lo que aumentará el promedio del grupo y el individual de manera artificial, lo que interfiere con la validez. Por otro lado, no discrimina entre estudiantes de bajo y alto desempeño.
- *Hacer trampa*. Este es uno de los retos principales de los profesores al aplicar exámenes, pues los estudiantes pueden hacer trampa de muchas formas y obtener ventaja de manera deshonesta: voltear a ver el examen de un compañero, utilizar un "acordeón", conocer las preguntas antes de hacer el examen, e incluso utilizar dispositivos digitales. La presencia de estas conductas genera falsos positivos, e incrementa la varianza sistemática en las puntuaciones.
- *Decisión indefendible de puntuación aprobatoria*. En general, en México los exámenes tienen una calificación mínima aprobatoria de 6.0. Si estos exámenes no son reproducibles o la distribución de las calificaciones es demasiado amplia (calificaciones demasiado altas y demasiado bajas), es difícil justificar por qué el 6.0 es la calificación de pase (Norcini, 2003).
- *Enseñar a la prueba ("teaching to the test")*. Cuando los alumnos han recibido preparación para contestar preguntas de un examen específico, como los exámenes de admisión a las licenciaturas o incluso los exámenes que elabora un profesor en particular, ello añade ventaja a algunos estudiantes y distorsiona el verdadero propósito de la enseñanza (Bond, 2008). Se ha visto que en ocasiones los estudiantes acceden a bancos de preguntas y prefieren aprenderlas de memoria en lugar de aprender los contenidos de los temas estudiados. Este factor es una amenaza a la fuente de evidencia de validez de relación con otras variables, porque no será posible generalizar los resultados de la prueba con el resto de los ítems del universo que miden el constructo deseado.
- *Habilidad para responder exámenes (testwiseness)*. Los alumnos que han presentado una gran cantidad de exámenes escritos a lo largo de su vida, se convierten en expertos para identificar la respuesta correcta en un ítem defectuoso. Por ejemplo, el hecho de que con frecuencia la opción correcta suele ser la más larga, la que tiene mayor detalle y está mejor escrita. Los alumnos que desarrollan estas habilidades como estrategias para responder pruebas, ocasionan que las respuestas no demuestren con veracidad el constructo de interés. Otro concepto relacionado es la "adivinanza educada" (*educated guessing)*, que consiste en que contestan por eliminación de las opciones menos probables, mas no porque realmente sepan la respuesta (Jurado-Núñez y Leenen, 2016).

CONFIABILIDAD

La confiabilidad es la característica de las evaluaciones que se refiere a que los puntajes sean consistentes de persona a persona, de instrumento a instrumento y de un conjunto de ítems a otro dentro del mismo universo de ítems (Cizek, 2009). Va de la mano de la validez en cuanto a que depende de la interpretación de las puntuaciones de la prueba, y forma parte de la fuente de evidencia de validez basada en la estructura interna (Tavakol y Dennick, 2011). Sin embargo, una prueba puede ser confiable, pero tener validez limitada. Tomemos como ejemplo un tiro al blanco con las marcas de los dardos, haciendo una analogía con las puntuaciones de una prueba de persona a persona, de instrumento a instrumento o de un conjunto de ítems a otro dentro del mismo universo de ítems (Figura 2).

Figura 2. Esquema de "tiro al blanco" para visualizar los conceptos de validez y confiabilidad



A. Prueba no confiable, sin validez.
B. Confiabilidad regular, validez regular.
C. Prueba confiable, pero sin validez.
D. Prueba muy confiable, pero sin validez.
E. Prueba confiable y con validez.

En el inciso A las marcas están muy dispersas en una prueba que no es confiable, porque no se concentran cerca del mismo sitio cada vez, ni dan en el centro (que es lo que se acerca más a lo que se desea evaluar). En el inciso B las marcas demuestran regular confiabilidad, pues no están tan dispersas, y validez regular porque, aunque algunas sí están en el centro, otras no. El inciso C demuestra una prueba confiable, pero con poca validez, ya que las marcas se concentran alrededor del mismo sitio, pero no en el centro, al igual que en el inciso D, aunque en este último están más cercanas entre sí. Finalmente, el inciso E demuestra lo deseable en cuanto a confiabilidad y validez: todas las marcas están concentradas en el centro.

¿Cómo se mide la confiabilidad? Los coeficientes de confiabilidad expresan la relación entre los puntajes que obtiene el mismo estudiante cuando presenta el mismo examen en dos ocasiones diferentes o en dos partes del mismo examen (Downing, 2004; Fraenkel et al., 2019). Esta relación da una idea de cuánta variación se puede esperar entre las diferentes

ocasiones, de persona a persona o de muestra de ítems a muestra de ítems. A continuación se describen algunas fórmulas para obtenerlos y adquirir información acerca de la consistencia interna de la prueba:

- *Kuder-Richardson*. Se utiliza para pruebas con variables dicotómicas (correcto/incorrecto), y otorga información sobre qué tan bien la prueba mide el constructo de interés (Kuder y Richardson, 1937). para conocer el valor de este coeficiente se pueden utilizar dos fórmulas:
  - KR20 – para pruebas con ítems con dificultad variable.

$$KR20 = \left(\frac{n}{n-1}\right)\left(\frac{1-\sum p \cdot q}{var}\right)$$

n= número de ítems

p= proporción de personas que aprueban el ítem

q= proporción de personas que reprueban el ítem

var= varianza para la prueba

  - KR21 – para pruebas con ítems con la misma dificultad.

$$KR21 = \left(\frac{n}{n-1}\right)\left(1-\frac{M(n-M)}{n \cdot var}\right)$$

n= número de ítems

M= media de la puntuación para la prueba

var= varianza de la prueba.

- *Alfa de Cronbach*. Es conocido como el coeficiente alfa, y se utiliza para pruebas con ítems con más de dos respuestas correctas, por ejemplo, preguntas con respuestas en forma de escala de valoración tipo Likert (Tavakol y Dennick, 2011).

$$\alpha = \frac{N\bar{c}}{\bar{v} + (N-1)\bar{c}}$$

$\propto$= alfa de Cronbach

N= número de ítems

$\bar{c}$= covarianza promedio inter-ítem

$\bar{v}$= varianza promedio

Los valores de los coeficientes de confiabilidad van de 0.00 a 1.00, donde 0 corresponde a una ausencia completa de relación y 1.00 el máximo posible de la relación (Fraenkel et al., 2019).

En el caso de pruebas de mediano impacto, si el valor es menor a 0.5, la consistencia interna es no aceptable; si es de 0.5 a 0.59, es pobre; si es de 0.6 a 0.69, es cuestionable; si es de 0.7 a 0.79, es aceptable; si es de 0.8 a 0.89, es buena; y mayor a 0.9 es excelente. Si la prueba es de alto impacto, el valor mínimo deseado es de 0.8 (Bland y Altman, 1997; Downing, 2004).

CONCLUSIONES

La validez es uno de los conceptos más importantes en evaluación del y para el aprendizaje. El modelo actual de validez es el resultado de una gran cantidad de investigaciones y discusiones entre expertos en el tema, y se le considera como un concepto holístico que se alimenta de diversas fuentes (contenido, proceso de respuesta, estructura interna, relación con otras variables y consecuencias). Validez se refiere a las inferencias que pueden hacerse con los resultados, más que a las pruebas o exámenes *per se*. Durante el proceso de evaluación deben cuidarse las amenazas a la validez, tanto de subrepresentación como de varianza irrelevante al constructo. La confiabilidad o reproducibilidad de los resultados, se integra como un elemento de fuente de evidencia de validez en el rubro de estructura interna.

REFERENCIAS

Abad, F. J., Olea, J., y Ponsoda, V. (2001). Analysis of the optimum number alternatives from the Item Response Theory. *Psicothema*, *13*(1), 152–158.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2018). *Estándares para pruebas educativas y psicológicas* (Original w). American Educational Research Association.

Bland, J., y Altman, D. (1997). Statistics notes: Cronbach's alpha.pdf. *British Medical Journal*, *314*, 572.

Bond, L. (2008). Teaching to the Test: Coaching or Corruption. *New Educator*, *4*(3), 216–223. https://doi.org/10.1080/15476880802234482

Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Carrillo-Avalos, B. A., Sánchez-Mendiola, M. y Leenen, I. (2020). El concepto moderno de validez y su uso en educación médica. *Revista de Investigación en Educación Médica*, *9*(33), 98–106. https://doi.org/https://doi.org/10.22201/facmed.20075057e.2020.33.19216

Carrillo-Avalos, B. A., Sánchez-Mendiola, M., y Leenen, I. (2020). Amenazas a la validez en evaluación: implicaciones en educación médica. *Investigación en Educación Médica*, *9*(34), 100–107. https://doi.org/10.22201/facmed.20075057e.2020.34.221

Chiavaroli, N. (2017). Negatively-worded multiple choice questions: An avoidable threat to validity. *Practical Assessment, Research and Evaluation*, *22*(3), 1–14.

Cizek, G. J. (2009). Reliability and validity of information about student achievement: Comparing large-scale and classroom testing contexts. *Theory into Practice*, *48*(1), 63–71. https://doi.org/10.1080/00405840802577627

Cook, D. A., y Hatala, R. (2016). Validation of educational assessments: a primer for simulation and beyond. *Advances in Simulation*, *1*(1), 1–12. https://doi.org/10.1186/s41077-016-0033-y

Cronbach, L. J., y Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

de Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical Education*, *44*(1), 109–117. https://doi.org/10.1111/j.1365-2923.2009.03425.x

Downing, S. M. (2003). Validity: On the meaningful interpretation of assessment data. *Medical Education*, *37*(9), 830–837. https://doi.org/10.1046/j.1365-2923.2003.01594.x

Downing, S. M. (2004). Reliability: on the reproducibility of assessment data. Medical education, *38*(9), 1006–1012. https://doi.org/10.1111/j.1365-2929.2004.01932.x

Downing, S. M., y Haladyna, T. M. (2004). Validity threats: Overcoming interference with proposed interpretations of assessment data. *Medical Education*, *38*(3), 327–333. https://doi.org/10.1046/j.1365-2923.2004.01777.x

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to Design and Evaluate Research in Education* (10th ed.). Mc Graw-Hill Education.

Haladyna, T. M., y Downing, S. M. (2004). Construct-Irrelevant Variance in High-Stakes Testing. *Educational Measurement: Issues and Practice*, *23*(1), 17–27. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-3992.2004.tb00149.x

Haladyna, T. M., Downing, S. M., y Rodriguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, *15*(3), 309–334. https://doi.org/10.1207/s15324818ame1503_5

Haladyna, T. M., Rodriguez, M. C., & Stevens, C. (2019). Are Multiple-choice Items Too Fat? *Applied Measurement in Education*, *32*(4), 350–364. https://doi.org/10.1080/08957347.2019.1660348

Hicks, N. A. (2011). Guidelines for identifying and revising culturally biased multiple-choice nursing examination items. *Nurse Educator*, *36*(6), 266–270. https://doi.org/10.1097/NNE.0b013e3182333fd2

Instituto Nacional para la Evaluación de la Educación. (2017). Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación. En *Diario Oficial de la Federación*. https://www.inee.edu.mx/wp-content/uploads/2019/04/P1E104.pdf

Jurado-Núñez, A., y Leenen, I. (2016). Reflexiones sobre adivinar en preguntas de opción múltiple y cómo afecta el resultado del examen. *Investigación en Educación Médica*, *5*(17), 55–63. https://doi.org/10.1016/j.riem.2015.07.004

Kane, M. T. (2009). Validating the interpretations and uses of test scores. In R. W. Lissitz (Ed.), *Validity: Revisions, New Directions and Applications* (pp. 39–64). Information Age Publishing, Inc.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160.

Lane, S. (2014). Evidencia de validez basada en las consecuencias del uso del test. *Psicothema*, *26*(1), 127–135. https://doi.org/10.7334/psicothema2013.258

Leenen, I. (2014). Virtudes y limitaciones de la teoría de respuesta al ítem para la evaluación educativa en las ciencias médicas. *Investigación en Educación Médica*, *3*(9), 40–55.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (pp. 13–103). MacMillan. https://doi.org/10.1002/j.2330-8516.1987.tb00244.x

Moreno, R., Martínez, R. J., y Muñiz, J. (2004). Directrices para la construcción de ítems de elección múltiple. *Psicothema*, *16*(3), 490–497. https://www.redalyc.org/articulo.oa?id=72716324

National Board of Medical Examiners. (2016). *Cómo elaborar preguntas para evaluaciones escritas en las áreas de ciencias básicas y clínicas*. *3*, 1–98.

Norcini, J. J. (2003). Setting standards on educational tests. *Medical Education*, *37*(5), 464–469. https://doi.org/10.1046/j.1365-2923.2003.01495.x

Padilla, J. L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, *26*(1), 136–144. https://doi.org/10.7334/psicothema2013.259

Rios, J., y Wells, C. (2014). Evidencia de validez basada en la estructura interna. *Psicothema*, *26*(1), 108–116. https://doi.org/10.7334/psicothema2013.260

Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice*, *24*(2), 3–13. https://doi.org/10.1111/j.1745-3992.2005.00006.x

Schuwirth, L. W. T., & van der Vleuten, C. P. M. (2011). General overview of the theories used in assessment: AMEE Guide No. 57. *Medical Teacher*, *33*(10), 783–797. https://doi.org/10.3109/0142159X.2011.611022

Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, *26*(1), 100–107. https://doi.org/10.7334/psicothema2013.256

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Capítulo 3

EVALUACIÓN DIAGNÓSTICA

Adrián Martínez González, Abigail Manzano Patiño,
Enrique Buzo Casanova, Melchor Sánchez Mendiola

"La finalidad de la evaluación reside en el análisis de la congruencia entre los objetivos y los logros."
Ralph W. Tyler, 1950.

INTRODUCCIÓN

La evaluación diagnóstica complementa y refuerza la evaluación continua que el profesorado lleva a cabo día con día en los espacios educativos. Constituye una fuente de información valiosa para los docentes, ya que les permite conocer de inicio cuáles son los aprendizajes logrados y aquellos que no se han consolidado, con el fin de reflexionar desde dónde partir en su proceso de enseñanza. En general, el estudiantado es sujeto de evaluación desde que ingresa al sistema educativo, particularmente en Educación Media Superior (EMS) o bachillerato y en Educación Superior (ES), realizando una gran cantidad de exámenes de todo tipo durante su formación. En tiempos recientes en los que la pandemia ha cambiado la forma convencional de interactuar entre los actores educativos, cabe reflexionar sobre los efectos que ha tenido en la educación y particularmente en la enseñanza dentro del aula.

La evaluación es un amplio campo multi e interdisciplinario, que en la actualidad ha tomado gran relevancia, lo que se refleja en el incremento de revistas internacionales indexadas que dedican sus espacios para publicar artículos exclusivamente de evaluación: *American Journal of Evaluation; Evaluation Review; Research Evaluation; Assessment and Evaluation in Higher Education; Educational Evaluation and Policy Analysis; Educational Research and Evaluation; Studies in Educational Evaluation*, *Revista Iberoamericana de Evaluación Educativa, Revista Electrónica de Investigación y Evaluación Educativa*, entre otras. Estos espacios académicos han impactado positivamente en el avance de conocimientos en evaluación educativa, aportando cada vez mayor sofisticación al campo de estudio, que hace necesario que tanto los profesionales de la evaluación como el cuerpo docente en general realicen esfuerzos para mantenerse actualizados en estos temas.

La evaluación también se considera un ejercicio profesional en la actualidad. Se han establecido incluso las competencias esenciales de los evaluadores profesionales (Stevahn, 2005), y se han desarrollado diversos programas de especialidad, maestría y doctorado en evaluación en varios países del mundo.

El objetivo de este capítulo es proveer un panorama general del proceso de evaluación diagnóstica en educación, algunas experiencias de aplicación de evaluaciones diagnósticas en la Universidad Nacional Autónoma de México (UNAM), así como recomendaciones para realizar dicha evaluación en estudiantes de EMS y ES.

EVALUACIÓN DIAGNÓSTICA: DEFINICIÓN Y APLICACIONES

En el primer capítulo de esta obra se describen diversas propuestas de definición del concepto de evaluación educativa, enfatizando la importancia de la evaluación del y para el aprendizaje, que implican un proceso sistemático y reflexivo de obtención de información cuantitativa y cualitativa sobre el aprendizaje del estudiantado, a través de la aplicación de diversos instrumentos (exámenes escritos u orales, exámenes prácticos, rúbricas, listas de cotejo, entre otros), para ser analizada con rigor metodológico, identificar fortalezas y áreas de oportunidad, emitir un juicio de su valía y así fundamentar la toma de decisiones sobre el proceso educativo orientado a su perfeccionamiento (capítulo 1 de esta obra).

Desde el punto de vista de su objetivo, la evaluación se puede clasificar en **diagnóstica,** sumativa y formativa (Downing S y Yudkowsky R, 2009; Sánchez-Mendiola M, 2018). En este capítulo nos enfocaremos en la **evaluación diagnóstica.** Este tipo de evaluación se realiza al principio de un curso o actividad académica, con la finalidad de determinar el nivel de conocimiento, habilidad o actitud del estudiante, e identificar si la persona evaluada cuenta con lo necesario para desarrollar de manera satisfactoria una actividad o proceso educativo (Arriaga M, 2015; Bombelli et al, 2012; Sánchez-Mendiola M, 2018).

En general la evaluación diagnóstica se realiza al inicio de un proceso de enseñanza y de aprendizaje, permite obtener juicios de valor para la toma de decisiones pedagógicas y fundamentar los planes educativos en concordancia con los saberes previos del estudiantado.

Constituye una herramienta de apoyo que favorece que el estudiante conozca sus alcances y limitaciones, y además le permita establecer estrategias de autorregulación para evaluar la calidad de su proceso de aprendizaje y acciones correctivas en la mejora de dicho proceso. El aprendizaje autorregulado es la capacidad que tiene el estudiante de seleccionar aquellas estrategias de regulación afectiva y emocional, en conjunto con ciertas actividades para mejorar su desempeño. Zimmerman (2002), destaca que los estudiantes que autorregulan su aprendizaje se caracterizan por participar activamente durante el mismo. Lo anterior, fomenta las habilidades metacognitivas y la influencia directa e indirecta de las emociones en la motivación y las actividades de aprendizaje. El entrenamiento adecuado de estas áreas favorece el proceso de aprendizaje y mejora el rendimiento (Ruiz, 2020).

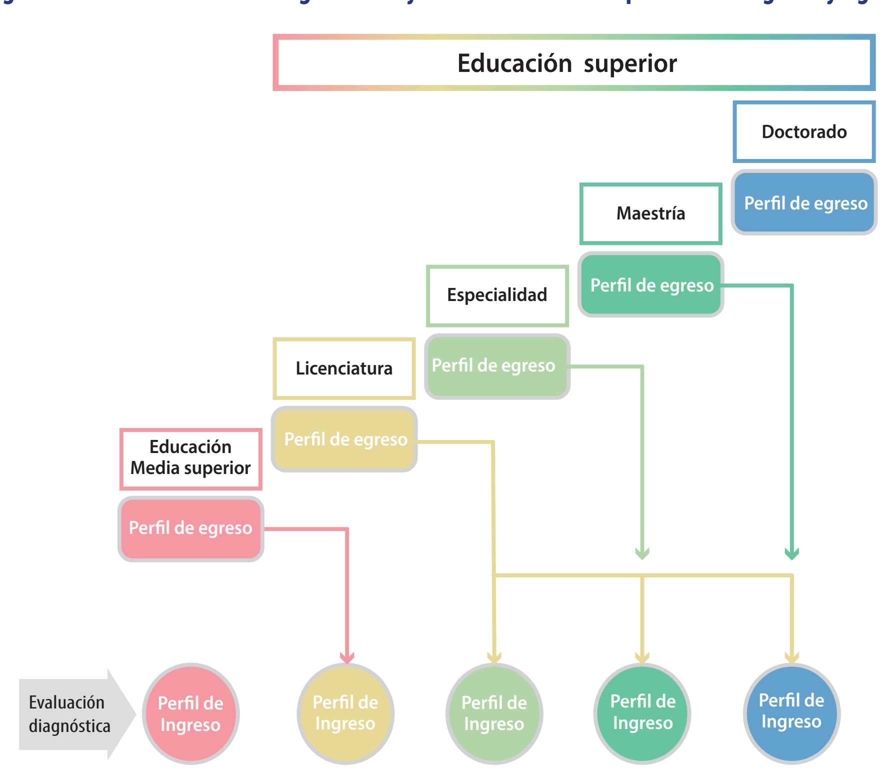
La evaluación diagnóstica inicial y la autoevaluación han sido destacadas por varios teóricos (Rosenshine et al.,1996). La acción que realiza el estudiante al cuestionarse sobre lo que

sabe en realidad y su desempeño actual, fomenta un proceso de autoevaluación centrándolo en el curso a realizar. El autoevaluarse y cuestionar su propio desempeño promueve la metacognición o autorregulación del aprendizaje, fomenta una mejor comprensión y análisis de los nuevos contenidos, al mismo tiempo que fomenta el trabajo colaborativo.

El proceso de evaluación diagnóstica en los espacios educativos se convierte entonces en una herramienta valiosa en el proceso de aprendizaje desde la etapa de planificación y evaluación del curso, debido a que provee de información necesaria sobre el nivel de conocimientos previos del estudiante. Es así que la evaluación diagnóstica como herramienta para la planificación permite mayor seguridad en la labor docente, al ser un predictor de los eventos del curso y brindando formalidad a la tarea académica.

La evaluación diagnóstica, al brindarnos información sobre los conocimientos previos que poseen los estudiantes, proporciona un panorama sobre los materiales de aprendizaje que deberían conocer y su nivel de dominio real. Para que sea efectiva, la evaluación debe contar con un perfil de referencia general, que incluya los conocimientos generales previos al curso en cuestión. De esta manera, el perfil de referencia que puede ser un plan de estudios, el programa de una materia o módulo, o el temario de un curso anterior, son el marco de referencia del que se parte en el instrumento, los reactivos que lo componen deben atender a los contenidos más relevantes que sirvan de soporte o base para el aprendizaje significativo, o lo nuevo que aprenderá el estudiante en el curso que inicia.

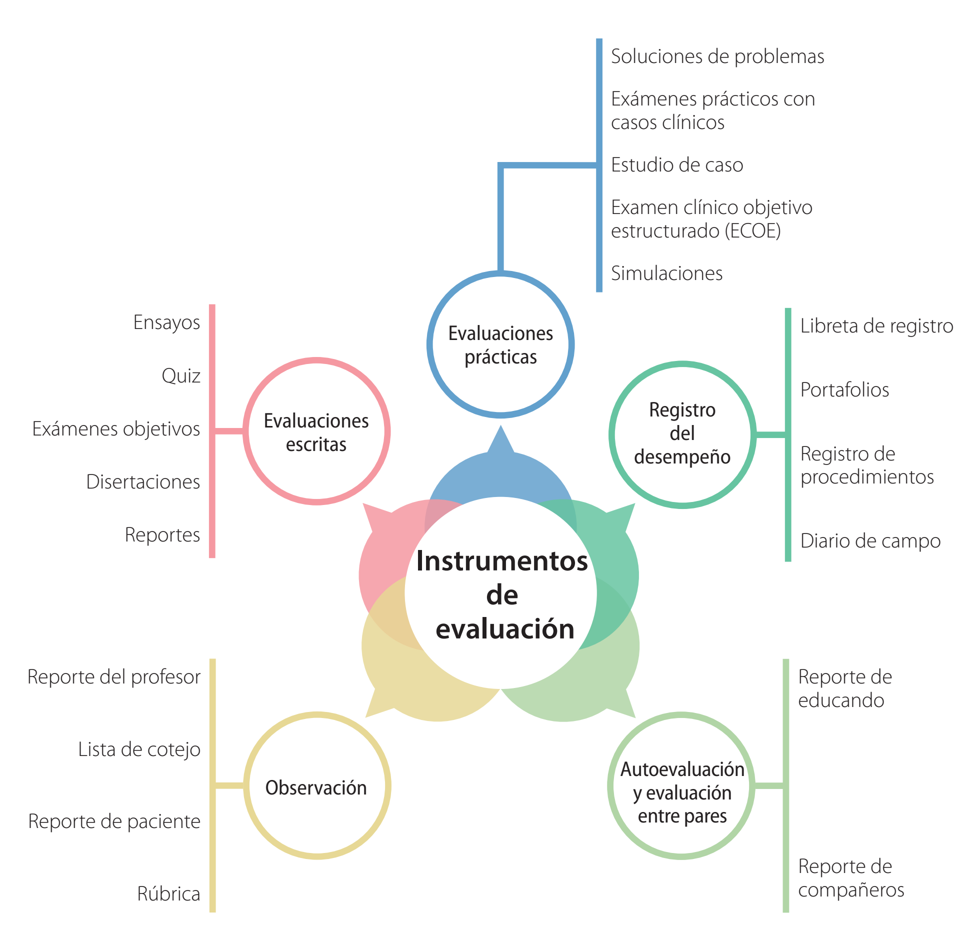
Figura 1. Las evaluaciones diagnósticas y su relación con los perfiles de ingreso y egreso



Esta información es de gran utilidad para el docente ya que le permite hacer adecuaciones en el contenido y en la implementación de las actividades académicas programadas, para corresponder a las características del alumno que participará en la actividad educativa. Debido a que la evaluación diagnóstica se encuentra ligada a perfiles (por ejemplo, los perfiles de ingreso o egreso a la EMS o a la ES) (Figura 1), las instituciones educativas también se ven fortalecidas con este tipo de evaluaciones, ya que les permite implementar estrategias enfocadas a mejorar el dominio de las competencias mínimas requeridas para iniciar en el nivel que se ha evaluado.

La evaluación diagnóstica no tiene límites muy precisos, debemos considerar que un examen de ingreso al bachillerato o licenciatura tiene un fuerte componente sumativo, pero también puede usarse como evaluación diagnóstica e incluso formativa si se provee la información a los docentes y estudiantes. De la misma forma, una evaluación diagnóstica puede funcionar como evaluación formativa, si se utiliza para dar retroalimentación a los estudiantes. La evaluación, al final del día, es un continuo de acciones de los diferentes actores del sistema, que intentan promover el aprendizaje significativo del estudiantado.

Figura 2. Tipos de instrumentos que pueden usarse en evaluación diagnóstica



Cuando los estudiantes ingresan a un nuevo ciclo o nivel educativo, las instituciones y los docentes se enfrentan al reto de determinar si los estudiantes tienen los conocimientos mínimos requeridos para continuar con su proceso de aprendizaje. Debido a que ningún estudiante aprende al mismo ritmo y la enseñanza dentro de los espacios educativos (aula, laboratorio, entre otros) también varía de profesor a profesor, la evaluación diagnóstica y los instrumentos para llevarla a cabo toman gran relevancia. Estos instrumentos son herramientas básicas que permiten obtener la información de lo que se quiere evaluar, y cada uno tiene ventajas y limitaciones para documentar el aprendizaje de los conocimientos, habilidades y destrezas (Downing, 2009; Shumway, 2003). Los instrumentos pueden clasificarse en las siguientes categorías (Figura 2):

- *Evaluaciones escritas*: ensayos, preguntas directas de respuesta corta, exámenes de opción múltiple, relación de columnas, disertaciones, reportes.
- *Evaluaciones prácticas*: exámenes prácticos con casos, examen clínico objetivo estructurado (ECOE).
- *Observación*: reporte del profesor, listas de cotejo, rúbricas, reporte de pacientes.
- *Portafolios y otros registros del desempeño*: libretas de registro, portafolios, diario de campo, registros de procedimientos.
- *Autoevaluación y evaluación por pares*: reporte del educando, reporte de los compañeros.

Cada uno de estos instrumentos tiene ventajas, desventajas, características psicométricas y recomendaciones para su implementación, que varían dependiendo del modelo educativo empleado, la normatividad institucional y el contexto. Los detalles y particularidades de uso se abordan en diferentes capítulos de este libro. El diseño, selección, implementación y acumulación de evidencia de validez de los instrumentos que se empleen para evaluar el aprendizaje de los estudiantes es responsabilidad de los profesores y de la organización evaluadora en las instituciones educativas (Brennan, 2006; Downing, 2009).

ALGUNAS EXPERIENCIAS DE EVALUACIONES DIAGNÓSTICAS EN LA DIRECCIÓN DE EVALUACIÓN EDUCATIVA (DEE) DE LA UNAM

La Universidad Nacional Autónoma de México tiene una amplia tradición en evaluación diagnóstica en los niveles de bachillerato y licenciatura, en donde se diseñan año con año exámenes diagnósticos que se aplican a todos los alumnos de primer ingreso al bachillerato y a la licenciatura (Valle R, 2012). Estos instrumentos de evaluación se diseñan siguiendo las buenas prácticas de evaluación para exámenes objetivos, estandarizados y a gran escala (Downing S y Haladyna T, 2006). El objetivo de dichas pruebas es evaluar los contenidos curriculares del nivel inmediato anterior, y aunque no se consideran de alto impacto ni tienen consecuencias para los estudiantes, son importantes para las facultades y escuelas. Los resultados permiten llevar a cabo actividades como: detección de estudiantes con bajo nivel académico a los que se apoya con cursos remediales, alumnos con alto nivel a los que se les

puede canalizar con grupos de investigadores, enviar resultados a los profesores, para que se puedan hacer adecuaciones en el contenido de sus actividades dentro de los espacios educativos. También se nutre a las entidades académicas con informes de resultados y listados de alumnos con información detallada de su desempeño en cada una de las asignaturas, módulos o cursos que componen los exámenes.

El examen diagnóstico que se aplica al bachillerato de la UNAM evalúa una población muy heterogénea, superior a 30,000 alumnos. Como los estudiantes proceden de cientos de planteles de enseñanza secundaria públicos y privados, para los subsistemas de EMS de la UNAM, la Escuela Nacional Preparatoria (ENP) y el Colegio de Ciencias y Humanidades (CCH), este examen representa el primer análisis sobre el desempeño de la población de nuevo ingreso. La prueba diagnóstica de bachillerato evalúa seis asignaturas: matemáticas, español, física, química, biología e historia, aunque matemáticas y español son las que tienen mayor peso en el examen, ya que alrededor de 34% de los reactivos valoran matemáticas y 18% español. Esto se debe a que estas dos asignaturas tienen mayor carga horaria en el plan de estudios que abarca la educación básica y por lo tanto se debe considerar en la estructura del examen. Algunos planteles de la UNAM como CCH Azcapotzalco o Naucalpan, han utilizado los resultados de estas evaluaciones para apoyar a los estudiantes con cursos o tutorías para subsanar las deficiencias en matemáticas y otras asignaturas durante el primer y segundo semestre. La DEE ha llevado a cabo, además, durante varios años, la aplicación del mismo examen diagnóstico a la población que egresa de la EMS de la UNAM a los tres años posteriores de su inicio, con el propósito de indagar si el paso por el bachillerato de la UNAM aportó un valor agregado a los conocimientos de los estudiantes. Un hallazgo documentado en esta evaluación, es que matemáticas es una de las asignaturas que mejores resultados ha mostrado, ya que, en todas las generaciones que se ha estudiado, esta registra uno de los mayores incrementos en la media del porcentaje de aciertos entre el ingreso y el egreso.

Los exámenes que se aplican a la población de primer ingreso a la licenciatura de la UNAM, también valoran a un importante número de estudiantes (más de 35,000). Para cada área de conocimiento (Físico matemáticas; Ciencias biológicas, químicas y de la salud; Ciencias sociales; Humanidades y artes) se aplica un examen que comprende ocho asignaturas: matemáticas, física, química, biología, historia universal y de México, literatura y geografía, por otra parte, el examen de filosofía se aplica solo en Humanidades y artes. Otro examen que se aplica a toda la población es el diagnóstico de español e inglés. Estos exámenes tienen doble finalidad, por un lado, los resultados que se les proporcionan a facultades y escuelas representan la primera descripción y análisis del desempeño alcanzado al egresar de EMS. Algunas entidades utilizan estos resultados, por ejemplo, las facultades de Medicina, Medicina Veterinaria y Zootecnia, la Facultad de Estudios Superiores Cuautitlán, y la Facultad de Artes y Diseño para diversos fines educativos. Debido a que alrededor del 80% de alumnos de primer ingreso proceden del bachillerato de la misma UNAM, estos resultados también benefician a la ENP y el CCH, ya que al darles a conocer como fue el desempeño de los estudiantes que egresan de sus planteles, les aporta evidencia para documentar si se alcanzaron las competencias curriculares esperadas en esta etapa,

necesarias para afrontar los retos que imponen los primeros semestres de la licenciatura que están por cursar.

Para la DEE, estos exámenes suponen un gran esfuerzo de diseño, logística y análisis que nutre a múltiples investigaciones sobre la población estudiantil. Cada año también se elabora una publicación con los resultados detallados de los exámenes, así como tablas dinámicas en Excel que los profesores o cualquier usuario pueden utilizar para conocer más sobre el desempeño de los estudiantes de nuevo ingreso al bachillerato y a la licenciatura. Los documentos en versión PDF y las tablas dinámicas en hoja de cálculo están disponibles en:

https://cuaieed.unam.mx/evaluacion_educativa#evaluaciones-bachillerato
https://cuaieed.unam.mx/evaluacion_educativa#evaluaci ones-licenciatura

Entre las publicaciones que se han derivado de estos exámenes podemos citar dos artículos recientes. "*Grado de conocimientos de los estudiantes al ingreso a la licenciatura y su asociación con el desempeño escolar y la eficiencia terminal. Modelo multivariado*" en el cual se estudió la asociación que tiene el examen diagnóstico con el desempeño escolar a lo largo de la licenciatura y la eficiencia terminal (egreso). Los resultados revelaron que, a mayor puntuación obtenida en el diagnóstico de ingreso, mejor desempeño escolar y mayor eficiencia terminal (Martínez et al., 2018). El segundo estudio indagó sobre las características que poseen los alumnos con mayor desempeño escolar en las licenciaturas del área de las ciencias biológicas, químicas y de la salud de la UNAM. Entre los hallazgos que se observaron destaca que, a mayor puntuación obtenida en las materias de biología, química, matemáticas, español e inglés de los exámenes diagnósticos de ingreso, mayor es el desempeño y el éxito académico en las licenciaturas de esta área *("Perfil del estudiante con éxito académico en las licenciaturas del área de las Ciencias Biológicas, Químicas y de la Salud*") (Martínez et al., 2020).

La pandemia por COVID-19 tuvo un impacto mundial en todos los ámbitos y la UNAM no quedó exenta de esta situación. El cierre de instalaciones a todos los niveles, supuso una modificación en la forma en que tradicionalmente se venía trabajando y enseñando en los espacios educativos. En la DEE, fue necesario desarrollar plataformas para realizar exámenes en línea, en particular EXAL, para contestar los exámenes diagnósticos de ingreso al bachillerato y la licenciatura en línea. Entre las ventajas que derivan de contestar los exámenes por medio de esta plataforma, se puede mencionar un ahorro importante en impresión de las pruebas, además de disminuir los retos logísticos que implicaban los exámenes presenciales supervisados. Fue más sencillo obtener la lectura de los resultados para cada estudiante, ya que no se tuvieron que leer las hojas de respuesta en papel, impactando en el tiempo que se destinaba a esta tarea. Adicionalmente, las facultades y escuelas no tuvieron que buscar y destinar espacios físicos para aplicar los exámenes, lo que requiere mayor logística y recurso humano. Es importante destacar que la plataforma EXAL posee una interfaz responsiva, lo que permite a los estudiantes responder el examen desde cualquier dispositivo con conexión a internet, facilitando su aplicación. Una desventaja que se observó es que como la plataforma permite que el estudiante conteste en cualquier horario, no fue posible

monitorear en tiempo real las horas invertidas en el proceso de respuesta del examen, aun cuando sí se pudieron registrar las veces que "abandonó" la pantalla del navegador y obtener el registro de hora de inicio y fin para cada estudiante, lo que permitió estimar el rango de tiempo aproximado que invirtió el estudiante para contestarlos.

La Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia (CUAIEED) desarrolló por medio de la Dirección de Evaluación Educativa (DEE) la Web App Pruéb@te UNAM Licenciatura y la Web App Pruéb@te UNAM Bachillerato que son herramientas de apoyo en la preparación de los exámenes de ingreso a la Educación Media Superior (bachillerato) y a la Educación Superior (Licenciatura) (https://www.pruebate.unam.mx). Su función es apoyar a los estudiantes por medio de exámenes muestra y repaso de los diferentes bloques que componen los exámenes, de esta manera, los aspirantes tienen un diagnóstico preciso sobre los temas y resultados de aprendizaje que dominan y los que necesitan preparar más. Los temarios, reactivos y ejercicios de los exámenes muestra están construidos con la misma metodología empleada en los exámenes reales, lo que le permite al estudiante tener una certeza en cuanto a su grado de conocimientos.

Estas herramientas han sido de gran apoyo para los estudiantes en su aprendizaje y preparación para exámenes sumativos de altas consecuencias, en un estudio reciente "Utilidad de la Plataforma "Pruéb@te UNAM Licenciatura" para Ingresar a la Educación Superior", se compararon los resultados obtenidos por los aspirantes usuarios de Pruéb@te UNAM para el examen de ingreso a la licenciatura en contraste con los no usuarios. Los hallazgos muestran que el rendimiento en el examen de selección de los aspirantes usuarios fue significativamente mayor comparado con los no usuarios, lo que se asoció con un incremento en la probabilidad de ser seleccionado (Sánchez-Mendiola, et al., 2021).

EVALUACIONES DIAGNÓSTICAS EN LA FACULTAD DE MEDICINA DE LA UNAM

Otro ejemplo de evaluaciones diagnósticas institucionales son las que se llevan a cabo en el área médica, por medio de exámenes teórico prácticos denominados ECOE (Evaluación Clínica Objetiva Estructurada), en los que se evalúa la capacidad clínica que tiene un médico en la práctica. Una de las aplicaciones de este tipo de examen, que se ha llevado a cabo en la Facultad de Medicina de la UNAM, ha sido con el propósito de evaluar un plan de estudios. En el artículo *"Evaluación diagnóstica de conocimientos y competencias en estudiantes de medicina al término del segundo año de la carrera: el reto de construir el avión mientras vuela"* (Martínez et al., 2014), se planteó la necesidad de evaluar el grado de conocimientos y competencias que los estudiantes poseen al terminar el segundo año de la licenciatura de medicina considerando un nuevo plan de estudios. Estos ejercicios toman relevancia por su valiosa contribución a los estudiantes, a los docentes y a la institución. Los resultados pueden aportan evidencia para valorar si los objetivos curriculares del plan de estudio fueron alcanzados, determinar qué campos de conocimiento se han logrado consolidar y aquellos que son deficientes, y al comunicar los resultados a los estudiantes, estos pueden ser capaces de buscar nuevas formas de aprender y convertir sus debilidades en áreas de oportunidad.

EN ESPACIOS DE APRENDIZAJE CON ALUMNOS

La evaluación diagnóstica al interior de los espacios de aprendizaje, supone una gran oportunidad para el docente, ya que al conocer los resultados de sus estudiantes tiene la oportunidad de reflexionar sobre su propia práctica, modificar su método de enseñanza y programar actividades de apoyo, ayudando así a disminuir el rezago en los contenidos académicos que más adelante dificulten el aprendizaje de otros temas de mayor complejidad. El docente puede complementar las evaluaciones diagnósticas institucionales como las que lleva a cabo la DEE de la UNAM, con otras realizadas específicamente al campo disciplinario en el que realiza el proceso educativo, de esa manera los resultados podrán ser más integrales y de mayor utilidad para los estudiantes y docentes.

La Comisión Nacional para la Mejora Continua de la Educación en México, dio a conocer en 2021 su plan de trabajo para realizar una evaluación diagnóstica en el nivel básico de enseñanza a nivel nacional (de segundo de primaria a tercero de secundaria), con el propósito de que, al volver a la modalidad presencial con la estabilización de la pandemia, las autoridades escolares y los profesores *"puedan identificar sus avances, los puntos de partida hacia la recuperación y los apoyos necesarios para fortalecer los aprendizajes"* en las áreas de lectura y matemáticas (https://www.mejoredu.gob.mx/images/publicaciones/Evaluacion_Diagnostica_AED.pdf). Considerando que este tipo de evaluaciones impacta positivamente a las instituciones, profesores y alumnos, estos ejercicios deben realizarse a nivel nacional, particularmente en EMS y ES.

CONCLUSIONES

Ante la interrogante de si los estudiantes alcanzaron o no los aprendizajes esperados ya sea porque comienzan un nuevo ciclo escolar, unidad, proyecto, o periodo, la evaluación diagnóstica aparece como una herramienta educativa poderosa.

Entre las recomendaciones que se plantean en este capítulo dirigidas al docente, está la de realizar evaluación diagnóstica específica al interior de su espacio de enseñanza que complemente la evaluación diagnóstica institucional, ya que esta combinación hará posible que se pueda desarrollar un método pedagógico que se ajuste a las necesidades de sus estudiantes. Algunas de las principales ventajas de las evaluaciones diagnósticas en educación son: permite adecuar los métodos y actividades de enseñanza; proporciona guías para intervenir de forma oportuna; facilita una retroalimentación personalizada y el acompañamiento del profesorado con los estudiantes; ayuda a reflexionar sobre la práctica a diferentes niveles (profesores, estudiantes, instituciones).

Las evaluaciones diagnósticas dejan de tener impacto si los resultados no se difunden entre los estudiantes, docentes y las autoridades institucionales o tomadores de decisiones. En la DEE se han llevado a cabo grandes esfuerzos para dar a conocer los resultados de estas evaluaciones diagnósticas tanto a autoridades como docentes, con el fin de que las utilicen en beneficio de sus estudiantes.

REFERENCIAS

Arriaga Hernández, M. (2015). El diagnóstico educativo, una importante herramienta para elevar la calidad de la educación en manos de los docentes. *Atenas*, *3*(31),63-74. [fecha de consulta: 21 de noviembre de 2021]. Disponible en: https://www.redalyc.org/articulo.oa?id=478047207007

Bombelli, EC, y Barberis, JG. (2012). Importancia de la Evaluación Diagnóstica en Asignaturas de Nivel Superior con Conocimiento Preuniversitario. *Revista Electrónica Gestión de las Personas y Tecnología*, *5*(13) [fecha de consulta: 21 de noviembre de 2021]. Disponible en: https://www.redalyc.org/articulo.oa?id=477847114004

Brennan, RL. (2006). *Educational Measurement. National Council on Measurement in Education*. Rowman & Littlefield Publishers. 4th Ed.

Downing, S. y Haladyna, T. (2006) Handbook of test development. USA: Lawrence Erlbaum Associates, Publishers. 793 pp.

Downing, SM, Yudkowsky, R. (2009). Introduction to Assessment in the Health Professions. En: Downing, SM, Yudkowsky (Eds). *Assessment in Health Professions Education*. New York, NY: Routledge. 1-21.

Herrera, J, y Ortega, E. (2022). Evaluación del y para el aprendizaje en educación universitaria. Clasificación de la evaluación de acuerdo con su finalidad, agente, impacto y referente [MOOC] https://www.coursera.org/lecture/evaluacion-educativa/clasificacion-de-la-evaluacion-de-acuerdo-con-su-finalidad-agente-impacto-y-ZNgMo

Martínez-González, A., Trejo-Mejía, J., Fortoul-van-der-Goes, T., Flores-Hernández, F., Sara Morales-López. S. y Sánchez-Mendiola. M. (2014). Evaluación diagnóstica de conocimientos y competencias en estudiantes de medicina al término del segundo año de la carrera: el reto de construir el avión mientras vuela. *Gaceta Médica de México*. *150*(1): 35-48. Disponible en: https://www.anmm.org.mx/GMM/2014/n1/GMM_150_2014_1_035-048.pdf

Martínez-González, A., Manzano-Patiño, A., García-Minjares, M., Herrera-Penilla, C., Buzo-Casanova, E. y Sánchez-Mendiola, M. (2018). Grado de conocimientos de los estudiantes al ingreso a la licenciatura y su asociación con el desempeño escolar y la eficiencia terminal. Modelo multivariado. *Revista de la Educación Superior*. *47*(188): 57-85. Disponible en: http://www.scielo.org.mx/pdf/resu/v47n188/0185-2760-resu-47-188-57.pdf

Martínez-González, A., Manzano-Patiño, A., García-Minjares, M., Herrera-Penilla, C., Buzo-Casanova, E. y Sánchez-Mendiola, M. (2020). Perfil del estudiante con éxito académico en las licenciaturas del área de las Ciencias Biológicas, Químicas y de la Salud. *Revista de la Educación Superior*. *49*(193): 129-152. Recuperado de http://resu.anuies.mx/ojs/index.php/resu/article/view/1029

Rosenshine, B. Meister, C. y Chapman, S. (1996). Teaching students to generate questions. A review of the intervention studies. *Review of Educational Research*, *66*(2), 181-221.

Ruiz M., H. (2020). ¿Cómo aprendemos? Una aproximación científica al aprendizaje y la enseñanza (1ª edición). Editorial Graó.

Sánchez Mendiola M. La evaluación del aprendizaje de los estudiantes: ¿es realmente tan complicada? Revista Digital Universitaria (RDU) 2018 noviembre-diciembre; *19*(6) ISSN: 1607-6079.

DOI: http://doi.org/10.22201/codeic.16076079e.2018.v19n6.a1

Sánchez-Mendiola, M., Rodríguez-Castillo S., Pérez-Herrera. N., García-Minjares M. y Martínez-González, A. (2021) Utilidad de la Plataforma “Prueb@te UNAM Licenciatura” para Ingresar a la Educación Superior. Revista de Investigación en Evaluación Educativa, *14*(2), Recuperado de https://revistas.uam.es/riee/article/view/riee2021_14_2_002

Stevahn, L, King, GG, y Minnema, J. (2005). Establishing essential competencies for evaluators. *American Journal of Evaluation*, *26*(1), 43-59.

Shumway JM, Harden RM. (2003). AMEE Guide No. 25: The assessment of learning outcomes for the competent and reflective physician. *Med Teach*. *25*(6):569-584.

Valle, F. (2012). El Sistema “Exámenes de diagnóstico y Autoevaluación y estudio de asignaturas del bachillerato de la UNAM”. *Revista Mexicana de Bachillerato a Distancia,* [S.l.], *4*(8). [fecha de consulta: 21 de noviembre de 2021]. Disponible en: <http://revistas.unam.mx/index.php/rmbd/article/view/44271/40022>. doi:http://dx.doi.org/10.22201/cuaed.20074751e.2012.8.44271.

Vera Vélez, Lamberto. (2013). Medición, “assessment” y evaluación del aprendizaje. Publicaciones Puertorriqueñas Editores, 6ª edición.

Zimmerman, B.J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory into Practice*, *41*(2), 64-72.

Capítulo 4

EVALUACIÓN FORMATIVA Y RETROALIMENTACIÓN DEL APRENDIZAJE

Tiburcio Moreno Olivos, Angélica Ramírez Elías

INTRODUCCIÓN

El origen del concepto evaluación formativa se situó a mediados de la década de 1960 en los Estados Unidos, fue Michael Scriven (1967) quien acuñó este concepto junto con el de evaluación sumativa. Los Estados Unidos son considerados como la cuna moderna de la evaluación (Nevo, 1998), de modo que la mayoría de los grandes teóricos de esta disciplina pertenecen a este país. Solo que, en aquella época, Scriven se refería a la evaluación formativa y sumativa del currículo. Posteriormente, dado el potencial que estos conceptos demostraron tener, se empezaron a emplear para referirse a la evaluación de los estudiantes, de suerte que en la actualidad se aplican prácticamente a todos los componentes del sistema educativo.

No obstante, Heritage (2010) señala que la evaluación formativa, en realidad, tuvo su desarrollo e impulso en un contexto distinto al de su origen (destacan los países de la Cuenca Asía-Pacífico e Inglaterra), como una reacción contra la visión dominante de la evaluación promovida por los Estados Unidos, considerada como sinónimo de medición y basada en la aplicación de pruebas estandarizadas, en lo que se conoce como el movimiento de estandarización de la evaluación.

Los pasados treinta años han visto un acelerado crecimiento en los estudios de la evaluación formativa. No obstante, aunque se habla con frecuencia de la evaluación formativa, se sabe poco acerca de esta y, quienes afirman emplearla en sus prácticas de evaluación, generalmente no lo hacen o al intentarlo desvirtúan su sentido original. Pese a tanto discurso, lo cierto es que la evaluación formativa continúa estando a la sombra de la evaluación sumativa, que es la evaluación que «realmente cuenta» en las escuelas. Pero, aún con este panorama, es innegable la importancia que la evaluación formativa tiene para el aprendizaje, al igual que ocurre con la retroalimentación como un componente fundamental de esta.

Este capítulo está integrado por dos apartados: el primero, refiere a la evaluación formativa y aborda cuestiones como: concepto, importancia de las interacciones, fuentes de evidencia y su calidad, propósitos de la evaluación y sus alcances; y el segundo, alude a la retroalimentación e incluye los siguientes temas: concepto, retroalimentación formativa, mejora de la calidad e impacto de la retroalimentación y el *feedforward*.

¿QUÉ ES LA EVALUACIÓN FORMATIVA?

Las decisiones acerca de la evaluación del estudiante no se toman de forma aislada, sino que son un reflejo del posicionamiento epistemológico que el evaluador adopta sobre el proceso de enseñanza-aprendizaje en su conjunto (Mellado, et al. 2021; Moreno, 2016). Como se sabe, la evaluación comprende varias dimensiones y cada una de ellas tiene un efecto significativo en qué se enseña y cómo se enseña, aunque en diferentes grados. Estas dimensiones responden a visiones pedagógicas dispares, que confrontan sobre el para qué se debe evaluar, qué evaluar y cómo hacerlo, quién evalúa, cuándo y qué hacer con los resultados de la evaluación.

Para Black y Wiliam (1998: 7-8) la evaluación formativa "abarca todas aquellas actividades realizadas por los profesores, y/o por sus estudiantes, que proporciona información que puede ser usada como retroalimentación para modificar las actividades de enseñanza y aprendizaje en las que ellos están comprometidos".

Para ilustrar lo dicho antes, veamos un ejemplo. Una profesora obtiene evidencia sobre el aprendizaje del estudiante en la materia de biología, pero esta evidencia no la adquiere empleando un «test» o examen escrito, sino mediante una investigación realizada cuidadosamente a través de la cual la profesora determina el estado actual del aprendizaje del estudiante, así como los siguientes pasos que se deben dar para avanzar en el aprendizaje mediante la acción pedagógica. Esta evidencia obtenida en la acción proporciona información que la docente usa como retroalimentación para el aprendizaje.

Para reunir evidencia que apoye el aprendizaje se destaca la importancia de la observación. Pero no se trata de una observación cualquiera sino de una observación «altamente cualificada», la cual puede darse en el escenario de las interacciones profesor-alumno, alumno-alumno, cuando se realizan las tareas y se desarrollan las acciones. En este tenor, Griffin (2007) argumentó que los humanos solo pueden proveer evidencia de aprendizaje cognitivo y afectivo a través de cuatro acciones observables: (1) lo que ellos dicen, (2) escriben, (3) crean (fabrican), o (4) hacen (ejecutan). Cualquiera que sea la fuente de evidencia, el papel del docente es construir o diseñar formas para obtener respuestas de los estudiantes que revelen su estado actual o real de aprendizaje (Sadler, 1989).

Una práctica fundamental de la evaluación formativa es la generación y recolección de información por parte del profesorado acerca de cómo se desarrolla el aprendizaje de los estudiantes mientras la enseñanza ocurre. Los docentes hacen un balance continuo del aprendizaje, centrando su atención en el desarrollo de la comprensión y las habilidades de los estudiantes a medida que el proceso de enseñanza-aprendizaje se lleva a cabo en tiempo real (Erickson, 2007: 179).

Se trata de una evaluación en la que el docente aprovecha cualquier momento de la clase (o incluso fuera de esta en otros espacios de la escuela) para extraer datos acerca del proceso de aprendizaje mediante las interacciones cotidianas con los estudiantes. Suele emplear técnicas e instrumentos informales de evaluación como observaciones, entrevistas, conversaciones espontáneas, etc., estos dispositivos informales se puede complementar con la aplicación de instrumentos formales de evaluación en momentos puntuales del curso. Un punto importante es que el docente lleve un registro y sistematización de la información recabada.

La evaluación formativa parte del entendido de que el aprendizaje y su evaluación son procesos complejos que no pueden reducirse a estándares, ya que se considera que el aprendizaje es fundamentalmente práctico, construido social e históricamente y donde el interés del estudiante por lo que aprende es fundamental.

Como se mencionó en líneas anteriores, la evaluación formativa es un proceso sistemático para obtener evidencia continua acerca del aprendizaje. Los datos reunidos son usados para identificar el nivel actual o real del estudiante y adaptar la enseñanza para ayudarle a alcanzar las metas de aprendizaje deseadas. En la evaluación formativa, los estudiantes son participantes activos con sus profesores, comparten metas de aprendizaje y comprenden cómo van progresando, cuáles son los pasos siguientes que necesitan dar, y cómo darlos (Heritage, 2007 en Moreno, 2016).

La evaluación formativa es un proceso de acompañamiento que conduce al desarrollo de las competencias del estudiante, el propósito es obtener información sobre su progreso de tal manera que se puedan identificar las necesidades de aprendizaje o limitaciones que presenta en el desarrollo de sus capacidades para implementar oportunamente una retroalimentación valiosa (UNESCO, 2021).

IMPORTANCIA DE LAS INTERACCIONES PARA LA EVALUACIÓN FORMATIVA

Dado que el propósito de aprendizaje exige una interacción dialógica –de los profesores con los estudiantes y de los estudiantes entre ellos– una tarea de aula debe diseñarse de tal modo que atraiga la atención de los estudiantes y los desafíe a reflexionar, expresar e intercambiar sus ideas acerca de la tarea. La interacción entre el profesor y los estudiantes ha sido caracterizada como una fuente principal de evidencia de la evaluación. Más específicamente, Black y Wiliam (2005) señalaron que las estrategias productivas de la evaluación formativa incluyen preguntas diseñadas por el profesor para explorar el aprendizaje de los estudiantes y generar momentos en los que ellos puedan intervenir y así promover el aprendizaje. Por su parte, Shavelson et al. (2008) sugirieron que para encontrar el vacío entre lo que los alumnos saben y lo que necesitan saber, el profesor debe desarrollar un conjunto de cuestiones centrales que lleguen al corazón de lo que se debe aprender en una clase específica. Además, el docente tiene que saber el momento correcto para preguntar estas cuestiones de modo que los estudiantes sean capaces de revelar lo que comprendieron, así como de aportar evidencia que pueda respaldar sus conocimientos. Al respecto, Chin (2007) ilustró el proceso de

preguntar y responder en el contexto de la ciencia: el profesor formula preguntas conceptuales para estimular las ideas de los estudiantes y ayudarlos a comprometerse con el pensamiento productivo, invita y alienta múltiples preguntas y respuestas, y sobre la marcha ofrece comentarios a sus respuestas.

En este sentido, Harlen (2007) señaló que no es solo el marco de las preguntas lo que importa sino también, la duración, particularmente el tiempo permitido para responderlas. Unas décadas antes, Rowe (1974) encontró que los profesores generalmente dan muy pocos segundos a los estudiantes para responder a las preguntas, lo que representa un tiempo muy corto aún para aquellas que solo requieren del recuerdo –y mucho menos para preguntas que demandan de los estudiantes proporcionar explicaciones o expresar sus ideas.

En un proyecto diseñado para mejorar el uso de la evaluación formativa de los profesores, los investigadores trabajaron con 24 docentes de matemáticas y ciencias y reunieron ideas para mejorar las técnicas de hacer preguntas y el tiempo de espera –la duración del silencio entre que un profesor hace una pregunta antes de hablar de nuevo si el estudiante no responde. Los profesores alteraron sus prácticas de hacer preguntas para incluir más tiempo de espera –por ejemplo, pidiendo a los alumnos que discutieran sus ideas en parejas antes de ser llamados al azar a responder (se instituyó una política de manos arriba). Además, los profesores no se refirieron a las respuestas como correctas o incorrectas, sino que pidieron a los estudiantes que explicaran las razones de sus respuestas, lo que les ofrecía una oportunidad para explorar las ideas que ellos tenían. El resultado claro de esta práctica fue que los docentes hicieron menos preguntas e invirtieron más tiempo en enseñar (Black, et. al., 2003).

Sobre este tema, Ruiz-Primo (2011) amplió sobre las preguntas como una fuente de evidencia para extender las secuencias de interacción referidas como conversaciones de evaluación. Desarrolló la noción de *diálogos de instrucción* como una estrategia pedagógica, las conversaciones de evaluación son concebidas como un diálogo que inserta a la evaluación como una actividad que ya está ocurriendo en el aula, que permite a los profesores profundizar en la naturaleza del pensamiento del estudiante y actuar pedagógicamente sobre esas perspectivas. Ruiz-Primo también señaló que la investigación sugiere que en las aulas donde los profesores frecuentemente participan en las conversaciones de evaluación, los estudiantes alcanzan niveles más altos de aprendizaje.

Aún en el contexto de la charla del estudiante, Harlen (2007) argumentó que los profesores pueden obtener conocimientos acerca del pensamiento de los estudiantes cuando preparan una situación en la que conversan unos con otros mientras el profesor «escucha a hurtadillas» sin participar en la conversación. Por supuesto, será esencial que la situación o actividad de aprendizaje diseñada esté bien estructurada y promueva entre los estudiantes diversas oportunidades para un intercambio fluido de ideas de modo que su pensamiento pueda ser revelado.

FUENTES DE EVIDENCIA Y SU CALIDAD

Con respecto a otras formas de obtención de evidencia con fines de evaluación, en el área de lectura, Bailey y Heritage (2008) ofrecieron una serie de estrategias para recabar información acerca de la lectura de los estudiantes, incluyendo lectura en voz alta, preguntas estratégicas centradas en el texto, y la solicitud de respuestas escritas acerca del texto como evidencia de su aprendizaje.

Como las funciones de maduración del educando no se desarrollan al unísono, la evaluación formativa es inevitablemente personalizada y los docentes necesitarán emplear estrategias que aprovechen el conocimiento individual que los estudiantes manifiestan. Cualquiera estrategia didáctica que un docente seleccione, debe cubrir el rango de estudiantes que se encuentra en el aula de modo que todos puedan mostrar donde se encuentran en su aprendizaje y tengan la expectativa de progresar desde su estado actual. Del mismo modo, las estrategias de evaluación formativa no deben incluir ningún elemento capaz de impedir que ciertos estudiantes puedan expresar cuál es su estado actual con relación a las metas, tales como el uso de un lenguaje incomprensible o imágenes que puedan resultar ofensivas para algunos de ellos.

Por otro lado, Erickson (2007: 189) introdujo la noción de que puede haber amenazas a la «formatividad» de la evaluación formativa. Este autor argumentó que para que la evaluación sea formativa debe ser oportuna y producir información que pueda orientar la práctica de enseñanza durante su desarrollo. Por esta razón, la inmediatez de la evidencia es un componente clave para la validez de la evaluación formativa.

En contraste con la evaluación estandarizada, la práctica de evaluación formativa descansa principalmente en los profesores. Además de determinar cómo y cuándo obtener evidencia, ellos también determinan quién evalúa. Por ejemplo, en una clase una profesora decide sobre unos estudiantes específicos lo que ella quiere evaluar en las interacciones individuales uno a uno. En otros casos, podría usar una estrategia de obtención de evidencia que se dedica a toda la clase o a grupos simultáneamente. En otro momento puede emplear la autoevaluación y la coevaluación.

PROPÓSITOS DE LA EVALUACIÓN

La evaluación tiene dos propósitos fundamentales: (1) proporcionar información sobre los actuales niveles de rendimiento de los estudiantes en el presente e (2) informar los pasos futuros que los profesores necesitan dar en las aulas para asegurar que los estudiantes progresan hacia los resultados deseados. Torrance y Pryor (2001) distinguieron entre dos tipos de evaluación. La primera se centra en tratar de descubrir *si* el aprendiz sabe, comprende, o puede hacer una cosa predeterminada y se caracteriza por tareas y cuestionarios cerrados o semi abiertos. La segunda pretende descubrir lo *que* el aprendiz sabe, comprende y puede hacer. Esta última se caracteriza por tareas y cuestionarios abiertos y está más orientada al desarrollo futuro que a la medición del rendimiento presente o pasado.

ALCANCES DE LA EVALUACIÓN FORMATIVA

La evaluación formativa tiene diversos alcances que se corresponden con distintos propósitos, entre los que se pueden identificar los siguientes:

a) Evalúa las dificultades de aprendizaje.
b) Retroalimenta a los actores del proceso educativo.
c) Reformula y reajusta las acciones del estudiante y docente.
d) Contribuye al desarrollo de los objetivos de aprendizaje.
e) Motiva al estudiante a la participación activa.
f) Informa y orienta sobre el progreso del proceso de aprendizaje.

Por supuesto, existen otros beneficios que la evaluación formativa puede aportar al proceso de enseñanza-aprendizaje, pero lo importante es reconocer que bien empleada tiene un enorme potencial para promover una evaluación sustentable y constructiva para los implicados en este proceso.

LA RETROALIMENTACIÓN

Uno de los componentes principales de la enseñanza y la evaluación formativa es la retroalimentación o *feedback*. De hecho, una de las funciones pedagógicas de la evaluación es precisamente la función retroalimentadora del aprendizaje. Nosotros consideramos que toda evaluación educativa, en *stricto sensu*, siempre será formativa y para lograrlo la función de retroalimentación es esencial.

Tal como ocurre con la evaluación formativa, en las escuelas se habla de la retroalimentación con cierta regularidad, pero se conoce menos de lo que habitualmente se piensa. Aunque es un concepto bastante familiar en el discurso pedagógico del profesorado universitario, la mayoría no sabe cómo aprovecharlo en beneficio del aprendizaje de sus estudiantes. Existen prácticas de evaluación del aprendizaje erróneas en el ámbito de la educación superior (Moreno, 2009). No toda retroalimentación es útil y valiosa para el aprendizaje, es más, existen prácticas que pueden ser contraproducentes o incluso nocivas para el aprendizaje escolar. Esto ocurre, por ejemplo, cuando se concibe a la retroalimentación de forma restringida constituyendo lo que se denomina como retroalimentación *tradicional* o *convencional*. Este tipo de retroalimentación se caracteriza como un acto de transmisión de información o de datos en un solo sentido (unidireccionalmente del profesor al estudiante), donde el receptor acepta pasivamente los comentarios de retroalimentación sin comprenderlos o sin saber qué hacer con ellos ni como emplearlos para mejorar su aprendizaje actual y futuro. Esta retroalimentación generalmente recibe un tratamiento superficial, es críptica, carente de un sustento o marco teórico y es concebida como un tipo de consejo o receta por parte del profesorado.

Aunque la retroalimentación se encuentra entre las principales influencias para el aprendizaje, el tipo de retroalimentación y la forma en que se ofrece pueden ser diferencialmente efectivas. En torno a este tema existen algunos asuntos típicamente delicados, incluido el

momento de la retroalimentación y los efectos de la retroalimentación positiva y negativa (Hattie y Timperley, 2007).

Como suele ocurrir con muchos de los conceptos empleados en el ámbito de la educación, la noción de retroalimentación ha sufrido cambios significativos durante el tiempo transcurrido desde sus orígenes hasta el presente, de modo que en la actualidad se conoce mejor el gran potencial que tiene para contribuir a la mejora del aprendizaje, así como las limitaciones y obstáculos de este proceso cuando no se realiza de forma correcta (Moreno, 2021).

En contraste con la retroalimentación tradicional o convencional, una perspectiva más reciente aboga por una *retroalimentación dialógica,* la cual consiste en establecer un diálogo abierto y constructivo entre el profesor y el estudiante, este último participa activamente mediante la formulación de preguntas, la solicitud de datos adicionales, la aclaración de dudas y el análisis de los comentarios que recibe, a fin de comprender dónde estuvieron los fallos y cómo puede utilizar esa información para mejorar su actuación o desempeño en próximas tareas (Carless, 2006; Moreno, 2021). Por supuesto, es posible que al inicio el estudiante muestre cierta resistencia a asumir ese papel activo que se le demanda, ante lo cual el docente tendrá que desplegar una serie de estrategias pedagógicas para incentivar la participación y el compromiso del estudiante con la retroalimentación, de suerte que gradualmente vaya adoptando un rol distinto en ese intercambio de información.

Un punto importante que es preciso mencionar es que la retroalimentación nunca opera en el vacío, para que tenga eficacia debe haber un contexto de aprendizaje al que se enfoca. La retroalimentación siempre debe basarse en algo, es poco útil cuando no existe un aprendizaje inicial o información firme. La retroalimentación no es más que parte del proceso de enseñanza y es lo que sucede después de que un estudiante ha respondido a la enseñanza inicial, cuando se proporciona información sobre algunos aspectos del rendimiento de la tarea del estudiante. La retroalimentación es más poderosa cuando aborda interpretaciones defectuosas, que cuando existe una nula comprensión, pues en este último caso, incluso puede ser amenazante para un estudiante: "Si el material estudiado es poco familiar o incomprensible, la retroalimentación tiene poco efecto sobre los criterios de rendimiento, ya que no hay forma de relacionar la nueva información con lo que ya se conoce" (Kulhavy 1977: 220, citado en Hattie y Timperley, 2007).

Para diversos investigadores, sin embargo, solo es posible hablar de retroalimentación cuando se cierra la brecha o vacío (entre lo que el estudiante sabe y lo que necesita saber) y tiene un impacto en el aprendizaje (Draper, 2009; Wiliam, 2011). La retroalimentación puede tener diferentes funciones dependiendo del entorno de aprendizaje, las necesidades del alumno, el propósito de la tarea y el paradigma de retroalimentación particular que se adopte (Knight y Yorke, 2003; Poulos y Mahony, 2008).

En el estudio de la retroalimentación se distinguen fundamentalmente dos paradigmas teóricos: el *cognitivista* y el *socio-constructivista*, y en la actualidad el énfasis está puesto en el segundo. La perspectiva cognitivista está estrechamente asociada con un enfoque directivo de «relato» en el que la retroalimentación se considera correctiva, y un experto proporciona

información a un receptor pasivo; mientras que en la perspectiva socio-constructivista se considera que la retroalimentación implica la provisión de comentarios y sugerencias para que los estudiantes puedan hacer sus propias revisiones y, a través del diálogo, ayudarlos a obtener nuevas comprensiones sin prescribir cuáles serán esas comprensiones (Archer, 2010). Un análisis más profundo de estas ideas nos permite afirmar que una perspectiva socio-constructivista enfatiza la naturaleza dinámica del aprendizaje donde el profesor también aprende del estudiante a través del diálogo y la participación en experiencias compartidas (Carless, et. al., 2011). No obstante, las dos perspectivas de retroalimentación antes descritas no son mutuamente excluyentes y deben ser vistas como refuerzos en lugar de como extremos opuestos de un *continuum*, cuando se considera la naturaleza precisa y el énfasis de la retroalimentación para apoyar las tareas, y las necesidades individuales y contextuales de los estudiantes.

En su artículo seminal, Sadler (1989) identificó a la retroalimentación del estudiante como el componente esencial de la evaluación formativa. Desde una perspectiva cibernética, adoptó una definición de retroalimentación de Ramaprasad, quien la consideró como la información acerca de la brecha entre el nivel actual y el nivel de referencia de un sistema parámetro que es usada para alterar la brecha en alguna forma. En consecuencia, Sadler (1989: 181) conceptualizó la evaluación formativa como una retroalimentación diseñada como circuito para cerrar la brecha entre el estado actual del estudiante y las metas deseadas. Para Sadler (1989: 120) un componente crítico para cerrar la brecha son los "juicios que los profesores hacen acerca de las respuestas del estudiante que pueden usarse para desarrollar y mejorar su competencia, rompiendo la aleatoriedad e ineficiencia del aprendizaje por ensayo y error". Estos planteamientos hacen eco de la obra de Pellegrino y Glaser (1982), quienes señalaron que la importancia primordial de la evaluación formativa es la exploración cuidadosa del profesor y el análisis del aprendizaje, lo cual puede conducir a ajustes sensibles en el aprendizaje individual de los estudiantes y las vías mediante las cuales progresarán. Por lo tanto, el propósito general de la obtención de evidencia de la evaluación formativa es permitir a los docentes responder al aprendizaje del estudiante a fin de mejorarlo, mientras el estudiante se encuentra participando en dicho proceso (Bell y Cowie, 2000).

RETROALIMENTACIÓN FORMATIVA

Como se comentó en líneas anteriores, la retroalimentación es un componente esencial de la evaluación formativa y, esta última, brinda información a los profesores y estudiantes acerca de qué están haciendo los educandos en relación a los objetivos de aprendizaje.

Ofrecer una buena retroalimentación es una de las habilidades que los profesores necesitan dominar como parte de una buena evaluación formativa. Otras habilidades de evaluación formativa incluyen tener claros los objetivos de aprendizaje, diseñar lecciones y tareas claras que comuniquen esos objetivos a los estudiantes, y –usualmente después de dar una buena retroalimentación– les ayuden a aprender como formular nuevos objetivos para sí mismos y planes de acción que conduzcan al logro de estos objetivos.

La retroalimentación puede ser muy poderosa si se hace bien. El poder de la evaluación formativa descansa en su enfoque de doble acercamiento, dirigido tanto a factores cognitivos como emocionales al mismo tiempo. La buena retroalimentación brinda a los estudiantes la información que necesitan para poder comprender dónde se encuentran en cuanto a su aprendizaje y qué tienen que hacer a continuación –el factor cognitivo. Una vez que entienden qué hacer y por qué, la mayoría de los estudiantes desarrolla la sensación de que tiene control sobre su propio aprendizaje: el factor motivacional. El desarrollo del aprendizaje autónomo en los estudiantes es una de las competencias que más se aprecian actualmente en los sistemas educativos en todo el mundo.

La evaluación formativa requiere una retroalimentación formativa. Esto quiere decir que no toda retroalimentación es formativa, esta puede incluso atentar contra el carácter formativo de la evaluación cuando, por ejemplo, se suministra a destiempo y está centrada en el yo como persona en lugar de dirigirse a la tarea o al proceso para realizar la tarea. Pero ¿cómo se define la retroalimentación formativa?

La retroalimentación formativa se define como la información que se comunica al estudiante con la intención de modificar su pensamiento o comportamiento con el fin de mejorar el aprendizaje. Según los expertos en esta temática la retroalimentación formativa debe ser multidimensional, no evaluativa, de apoyo, oportuna, específica, creíble, infrecuente y genuina (Shute, 2008). La retroalimentación formativa generalmente se presenta como información a un alumno en respuesta a alguna acción de su parte. Se presenta en una variedad de tipos (por ejemplo, verificación de la precisión de la respuesta, explicación de la respuesta correcta, sugerencias, ejemplos prácticos) y se puede administrar en varios momentos durante el proceso de aprendizaje (por ejemplo, inmediatamente después de una respuesta, después de algún tiempo transcurrido).

MEJORA DE LA CALIDAD E IMPACTO DE LA RETROALIMENTACIÓN

El aprendizaje sin retroalimentación es como un tiro con arco a ciegas: simplemente no es posible cumplir con un determinado estándar o norma si no sabes qué tan bien lo estás haciendo. Por lo tanto, la retroalimentación permite acertar al objetivo de manera consistente al proporcionar información que se puede utilizar para mejorar el rendimiento. Pero la retroalimentación y la orientación expertas también pueden mejorar y acelerar lo que se logra: pueden llevar a un estudiante a un nivel de rendimiento mucho más alto, y en algunos casos más rápidamente, de lo que hubiera sido factible. Dicho de otra manera, sin una retroalimentación adecuada, es probable que los estudiantes tengan un rendimiento bajo, sin poder demostrar todo lo que son capaces de lograr. Un compromiso con los altos estándares y la calidad de la retroalimentación van como «la mano al guante».

Tradicionalmente, la retroalimentación en educación superior ha sido proporcionada por el profesor, en forma de una calificación o nota acompañada de escuetos comentarios (generalmente escritos) sobre la tarea o evaluación de un estudiante. La calificación proporciona un juicio sumario sobre qué tan bien lo ha hecho un estudiante, mientras que los

comentarios cumplen una función de diagnóstico, y describen cómo se llegó a ese juicio, destacando las fortalezas y debilidades, o evaluando la calidad del trabajo del estudiante en relación con cada uno de los criterios de evaluación empleados. Por lo general, la responsabilidad recae en el estudiante para remediar cualquier deficiencia o debilidad identificada, en una oportunidad futura que pueda haber para realizar una tarea o evaluación similar.

En la educación superior del siglo XXI, este enfoque continúa siendo ampliamente utilizado, pero está experimentando cada vez más modificaciones significativas o está siendo complementado con otros enfoques, a medida que los docentes universitarios se esfuerzan por encontrar formas de maximizar el impacto de la retroalimentación. Apoyar todas estas iniciativas es una preocupación compartida de que la retroalimentación debería «hacer una diferencia», que refleja el influyente argumento presentado por Paul Black y sus colegas de que la retroalimentación solo puede servir para aprender completamente "si se trata tanto de evocar la evidencia como de una respuesta a esa evidencia usándola de alguna manera para mejorar el aprendizaje" (Black *et al*. 2003: 122).

FEEDFORWARD: IR MÁS ALLÁ DE LA RETROALIMENTACIÓN

Otro concepto que ha surgido recientemente en el estudio de esta temática es el de *feedforward* (Hounsell *et al.* 2007: 4-5). Se trata de un proceso que consiste en aumentar el valor de la retroalimentación para el estudiante enfocando los comentarios no solo en el pasado y el presente (lo que se escribió o demostró en el trabajo bajo escrutinio) sino también en el futuro: lo que los estudiantes podrían aspirar a hacer, o a hacerlo de manera diferente, en la siguiente tarea o evaluación, si han de seguir haciéndolo bien o mejor. Un análisis de los comentarios de los tutores en dos módulos de Economía (Johnston *et al.* 2008) proporcionaron evidencias que respaldan la necesidad de un cambio en el énfasis: menos del 1% de los comentarios se centraron en el trabajo futuro de la materia.

Una segunda modificación de la retroalimentación tradicional apunta a transformar el feedforward, pero a través de «trabajos de curso acumulativos» (Hounsell *et al.* 2007: 6-7). Esto implica construir en el proceso de retroalimentación una oportunidad más inmediata para que el alumno haga uso de la retroalimentación, a diferencia de lo que ocurre en la educación superior masiva actual (Hounsell y Hounsell 2007). La forma que suele adoptar es el ciclo borrador-comentario-revisión-reenvío, es decir, se da retroalimentación sobre un borrador, plan, esquema o extracto, que luego el estudiante vuelve a trabajar a la luz de los comentarios realizados antes de volver a enviarlo (Hounsell *et al.* 2007, para ver ejemplos de varias áreas temáticas).

Un inconveniente potencial de este enfoque es que a menos que se planifique cuidadosamente, se corre el riesgo de aumentar enormemente la carga de trabajo de los docentes. Una forma de eludir este peligro es la compensación directa, en la que el *feedforward* sustituye y reemplaza a la retroalimentación. Por lo tanto, los comentarios se dan antes de la presentación, pero no después, lo que se comunica al alumno sobre el trabajo terminado es una calificación o nota, y quizás también comentarios colectivos o de «toda la clase» (donde

los comentarios individuales a cada alumno se reemplazan por un conjunto de comentarios único y más sencillo dirigido al grupo o clase en general).

Otra forma de avanzar es rediseñar las tareas, dividiéndolas en pasos más pequeños que se vinculan entre sí y se construyen unos sobre otros. Los pasos sucesivos se intercalan con comentarios de «feedforward», pero en forma menos costosa, a modo de comentarios para toda la clase en lugar de los comentarios individuales (vea McCreery 2005, para una explicación de cómo se logró esto en un gran curso de Historia de primer año en la Universidad de Sydney mediante la reconfiguración de las tutorías; ver ejemplos de una variedad de otras áreas temáticas en Hounsell *et al*. 2007: 4 y siguientes).

Para lograr una retroalimentación más rápida, se propone una tercera modificación que apunta a aumentar el impacto de la retroalimentación al acelerar su provisión e inmediatez para los estudiantes, y así ayudar más directamente a su desempeño posterior en las evaluaciones sumativas. Es más común encontrarlo en cursos más grandes donde las preguntas de opción múltiple o similares son un componente importante en la mezcla de evaluación general. En tales casos, por lo general toma la forma de un recurso computarizado en línea que permite a los alumnos en varios momentos de un curso probar su comprensión y obtener retroalimentación constructiva sobre aquellos ítems que responden incorrectamente. Sin embargo, las formas de baja tecnología de retroalimentación más rápida también son factibles al vincular la retroalimentación rápida de toda la clase a las actividades de tutoría.

Existen nuevas oportunidades de retroalimentación. Un segundo ejemplo, pero muy diferente, extiende la retroalimentación a los exámenes, donde en muchos cursos y materias (por supuesto, no en todos), los alumnos a menudo solo reciben calificaciones generales o notas como un indicador de lo bien que lo han hecho. En términos generales, las cargas de calificación y los tiempos de entrega de los resultados de los exámenes hacen que no sea práctico ofrecer a los estudiantes el tipo de comentarios individualizados que normalmente reciben en los trabajos de curso, pero la retroalimentación se puede proporcionar de otras dos formas. Una es a través de retroalimentación colectiva o de toda la clase, por ejemplo, algunas observaciones generales sobre las fortalezas y limitaciones evidentes en las respuestas de los estudiantes en un examen determinado o en un grupo de preguntas del examen. Otra es bajo el disfraz del *feedforward* (Hounsell 2007) donde, además de permitir que los estudiantes tengan acceso a exámenes anteriores, el profesor también proporciona un comentario en forma de nota sobre cuáles fueron y cómo se pudieron abordar las preguntas anteriores, qué trampas potenciales podría estar a la espera de los incautos, y las ventajas y limitaciones comparativas de diferentes enfoques para una pregunta en particular (Hounsell *et al.* 2006). El *feedforward* puede considerarse un tipo de *evaluación formativa preventiva* (Carless 2007), en el sentido de involucrar acciones para prevenir el potencial de conceptos erróneos que, de otro modo, podrían limitar lo que los estudiantes pudieran lograr.

REFLEXIONES FINALES

En este capítulo hemos abordado dos temas que están íntimamente relacionados como son la evaluación formativa y la retroalimentación. Estos dos elementos juegan un papel crucial en las posibilidades (o no) que tengan los estudiantes de adquirir un aprendizaje profundo, relevante y significativo. Pero esta meta deseable no ocurrirá de forma espontánea, solo con ponerlos en marcha en el aula, de poco servirá su empleo si se hace de forma errática o torpe, se requiere que la evaluación sea cuidadosamente planeada e implementada, pues de no hacerlo así, se corre el riesgo de que no solo no logre su cometido principal: que es promover el aprendizaje, sino que acabe convirtiéndose en un obstáculo para la adquisición y el desarrollo del aprendizaje de los estudiantes. De ahí la necesidad de prestar atención al desarrollo de las competencias evaluadoras del profesorado universitario, es vital que los docentes sepan crear y llevar a cabo buenas evaluaciones de aula.

La evaluación formativa es diseñada para proveer retroalimentación o *feedback* al menos en dos niveles. Primero, brinda retroalimentación al profesor acerca de los niveles actuales de comprensión del estudiante, y segundo, informa cuáles son los pasos siguientes que deben darse en el aprendizaje. La retroalimentación es crucial para guiar al estudiante en sus próximos pasos y permite establecer un lazo entre él y el profesor, que se ven implicados en un proceso continuo de intercambio de información y orientación. Para ello, el cuidado en la recogida y el análisis de los datos de evaluación es esencial. Como hemos visto, las interacciones cotidianas y las preguntas formuladas a los estudiantes son un medio clave para obtener las evidencias.

Una retroalimentación efectiva, por parte de los profesores, brinda información clara, descriptiva, y basada en criterios. Con ello, se indica a los estudiantes: dónde están ubicados en una progresión de aprendizaje, cómo difiere su comprensión de la meta de aprendizaje deseada, y cómo pueden avanzar hacia esa meta. El profesor ejecuta los pasos para cerrar el vacío entre el aprendizaje actual del estudiante y la meta esperada, modifica la enseñanza, evalúa nuevamente para dar información adicional acerca del aprendizaje, modifica nuevamente la enseñanza, etcétera.

En la evaluación formativa, los estudiantes deben ser capaces de usar la retroalimentación para mejorar su aprendizaje, pero esa mejora del aprendizaje depende de la participación activa de ellos en su propia evaluación. En la evaluación formativa, los alumnos desarrollan habilidades de autoevaluación y evaluación de pares, de modo que, como Sadler (1989) sugiere, ellos colaboran con sus profesores en desarrollar una comprensión compartida del estado actual de su aprendizaje, y lo que necesitan hacer para avanzar en él. Algunos de los beneficios para los estudiantes que conlleva esta práctica son: a) reflexionan sobre su aprendizaje, monitorean lo que saben y comprenden y determinan cuando necesitan más información; b) desarrollan estrategias de autorregulación, y son capaces de adaptar sus estrategias para satisfacer sus propias necesidades de aprendizaje; c) colaboran con sus profesores para determinar los criterios de éxito para cada etapa a lo largo de la progresión del aprendizaje (Moreno, 2016).

La buena retroalimentación contiene información que un estudiante puede usar, lo que significa que este tiene que poder escucharla y entenderla. Los estudiantes no pueden escuchar

algo que está más allá de su comprensión; tampoco pueden escuchar algo si no están escuchando o sienten que no vale la pena escuchar. Por ello, una buena retroalimentación debe ser parte de un entorno de evaluación en el que los estudiantes vean la crítica constructiva como algo positivo y comprendan que el aprendizaje no puede ocurrir sin la práctica. No es razonable que a los educandos se les brinde retroalimentación y después no se les ofrezcan oportunidades para usarla. Tampoco es válido que el docente formule a los estudiantes una crítica constructiva y luego emplee esa crítica en su contra en una calificación o evaluación final.

REFERENCIAS

Archer, J. C. (2010). State of the science in health professional education: Effective feedback, *Medical Education*, *44*(1), 101-108. https://doi.org/10.1111/j.1365-2923.2009.03546.x

Bailey, A. y Heritage, M. (2008). *Formative assessment for literacy, grades K-6*. Thousand Oaks, CA: Corwin Press.

Bell, B. y Cowie, B. (2000). The characteristics of formative assessment in science education, *Science Education, 85*, 536-553.

Black, P. y Wiliam, D. (1998). Assessment and classroom learning, *Assessment in Education: Principles, Policy and Practice*, *5*(1),7-74.

Black, P. y Wiliam, D. (2005). Changing teaching through formative assessment: Research and practice, En: *Formative Assessment: Improving learning in secondary Classrooms* (pp.223-240). Centre for Educational Research and Innovation (CERI). Paris: Organisation for Economic Cooperation and Development.

Black, P., Harrison, C., Lee, C., Marshall, B. y Wiliam, D. (2003). *Assessment for Learning: Putting It into Practice*. Maidenhead: Open University Press.

Carless, D. (2006). Differing perceptions in the feedback process, *Studies in Higher Education*, *31*(2), 219-33.

Carless, D. (2007). Conceptualizing pre-emptive formative assessment, *Assessment in Education*, *14*(2), 171-184.

Carless, D., Salter, D., Yang, M. y Lam, J. (2011). Developing sustainable feedback practices, *Studies in Higher Education*, *36*(5), 395-407.

Chin, C. (2007). Teacher questioning in science classrooms: Approaches that stimulate productive thinking, *Journal of Research in Science Teaching, 44*, 815-843.

Draper, S. (2009). What are learners actually regulating when given feedback? *British Journal of Educational Technology, 40,* 306-315. https://doi.org/10.1111/j.1365-2923.2009.03546.x

Erickson, F. (2007). Some thoughts on "proximal" formative assessment of student learning, *Yearbook of the National Society for the study of Education*, *106*, 186-216.

Griffin, P. (2007). The comfort of competence and the uncertainty of assessment, *Studies in Educational Evaluation, 33*, 87-99.

Harlen, W. (2007). Formative classroom assessment in science and mathematics En: J. H. McMillan (Ed.), *Formative classroom assessment: Theory into Practice* (pp. 116-135). New York: Teachers College Press.

Hattie, J. y Timperley, H. (2007). The power of feedback, *Review of Educational Research*, *77*, 81-112. https://doi.org/10.3102/003465430298487

Heritage, M. (2010). Formative Assessment and Next-Generation Assessment Systems: Are We Losing an Opportunity? Artículo presentado en el *Council of Chief State School Officers*, Washington, septiembre.

Hounsell, D. y Hounsell, J. (2007). Teaching-learning environments in contemporary mass higher education. In: N. J. Entwistle, P. Tomlinson y J. Dockrell (Eds.), *Student Learning and University Teaching (Psychological Aspects of Education-Current Trends). British Journal of Educational Psychology Monograph,* Serie II, No. 4. Leicester: British Psychological Society, 91-111.

Hounsell, D., McCune, V., Hounsell, J. y Litjens, J. (2006). Investigating and enhancing the quality of guidance and feedback to undergraduate students. Artículo presentado en la *Third Biennial Northumbria/EARLI SIG Assessment Conference*, Northumbria, 30 Agosto -1 Septiembre. Disponible en: http://www.etl.tla.ed.ac.uk/docs/earlinth06.pdf

Hounsell, D., Xu, R. y Tai, C. M. (2007). *Balancing Assessment of and Assessment for Learning (Scottish Enhancement Themes: Guides to Integrative Assessment, no.2)*. Gloucester: Quality Assurance Agency for Higher Education. Disponible en: https://www.enhancementthemes.ac.uk/docs/ethemes/integrative-assessment/guide-no-2---balancing-assessment-of-and-assessment-for-learning.pdf?sfvrsn=8e2cf981_10

Johnston, C., Cazaly, C. y Olekalns, N. (2008). The first-year experience: perceptions of feedback. Artículo presentado en *Universitas 21 Conference on Teaching and Learning, Does Teaching and Learning Translate? Learning Across the U21 Network*, University of Glasgow, Febrero.

Knight, P. T. y Yorke, M. (2003). Assessment, learning and employability. Maidenhead, UK: SRHE/Open University Press.

McCreery, C. (2005). Less is more: rethinking assessment in a first-year history unit, *Synergy*, *22*, 23-26. Disponible en: https://educational-innovation.sydney.edu.au/synergy/pdfs/2223.pdf

Mellado, P.C., Sánchez, P. y Blanco, M. (2021). Tendencias de la evaluación formativa y sumativa del alumnado en Web of Sciences. *ALTERIDAD. Revista de Educación*, *16*(2), 170-183. https://doi.org/10.17163/alt.v16n2.2021.01

Moreno, T. (2009). La evaluación del aprendizaje en la universidad: tensiones, contradicciones y desafíos, *Revista Mexicana de Investigación Educativa*, *XIV*(41), 563-591.

Moreno, T. (2016). *Evaluación del aprendizaje y para el aprendizaje. Reinventar la evaluación en el aula*. Universidad Autónoma Metropolitana-Cuajimalpa. https://bit.ly/2RZYAL2

Moreno, T. (2021). *La retroalimentación. Un proceso clave para la enseñanza y la evaluación formativa*. Universidad Autónoma Metropolitana-Cuajimalpa. https://bit.ly/3wuCDCG

Nevo, D. (1998). La evaluación mediante el diálogo: una contribución posible al perfeccionamiento de la escuela, *Perspectivas, XXVIII*(1), 87-100.

Pellegrino, J. W. y Glaser, R. (1982). Analyzing aptitudes for learning: Inductive reasoning, En: R. Glaser (Ed.), *Advances in instructional Psychology,* 2, pp. 269-345). Hillsdale, NJ: Lawrence Erlbaum.

Poulos, A. y Mahony, M. J. (2008). Effectiveness of feedback: The students' perspective, *Assessment & Evaluation in Higher Education*, *33*, 143-154. https://doi.org/10.1080/02602930601127869

Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instructional dialogues in assessing students for science learning. Special issues in assessment for learning, *Studies of Educational Evaluation, 37*(1), 15-24.

Rowe, M. B. (1974). Wait time and rewards as instructional variables, their influence in language, logic, and fate control: Part II, rewards, *Journal of Research in Science Teaching, 11*, 291-308.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems, *Instructional Science*, *18*, 119-144.

Scriven, M. (1967). The methodology of evaluation, en *Perspectives of curriculum evaluation,* editado por Ralph Tyler, Robert Gagné, y Michael Scriven. Chicago: Rand McNally.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A. et al. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers, *Applied measurement in Education*, *21*(4), 295-314.

Shute, V. J. (2008). Focus on formative feedback, *Review of Educational Research 78*, 153–189. https://doi.org/10.3102/0034654307313795

Torrance, H. y Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory*, British Educational Research Journal, 27*, 615-631.

UNESCO (2021). Evaluación formativa: Una oportunidad para transformar la educación en tiempos de pandemia. Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE).

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3-14. https://doi.org/10.1016/j.stueduc.2011.03.001

Capítulo 5

EVALUACIÓN SUMATIVA Y EXÁMENES DE ALTO IMPACTO

Melchor Sánchez Mendiola, Laura Delgado Maldonado

"...el uso de la puntuación de un examen definitivamente implica consecuencias; de otra manera uso es solo una abstracción."
Robert L. Brennan

"No todo lo que puede ser contado cuenta, y no todo lo que cuenta puede ser contado."
Albert Einstein

EVALUACIÓN SUMATIVA

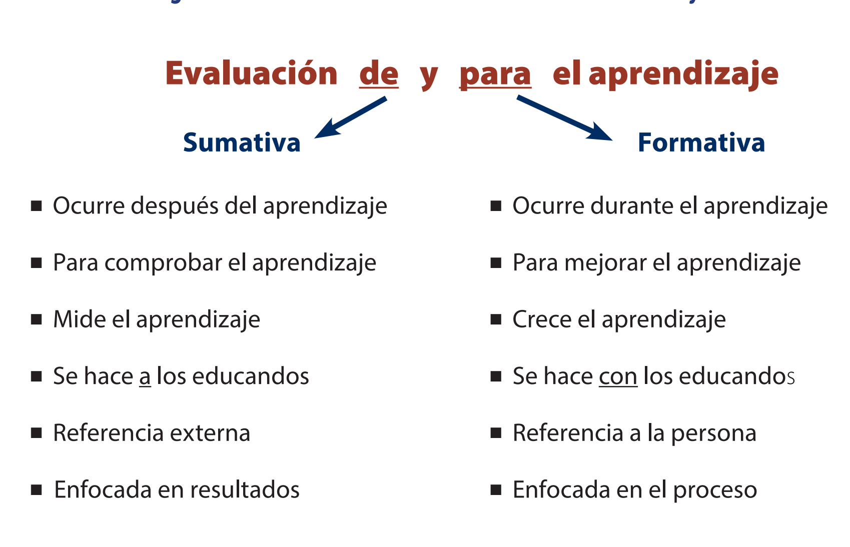
Como se ha descrito en otros capítulos de este libro, la evaluación puede clasificarse por su objetivo en diagnóstica, formativa y sumativa (capítulo 1). En este capítulo se describirán los principales conceptos de evaluación sumativa y de exámenes de alto impacto (EAI), también llamados de altas consecuencias (AERA, APA y NCME, 2014; INEE, 2017). La **evaluación sumativa** es aquella que se utiliza para determinar el aprendizaje del estudiante, adquisición de habilidades y logro académico al concluir una actividad instruccional específica, que puede ser el final de una unidad, curso, programa, año académico o semestre (Bennett, 2015; Miller et al, 2013; UNESCO, 2021). Este tipo de evaluación está compuesto por la suma de valoraciones efectuadas durante un curso, para determinar al final del mismo el nivel con que los objetivos de la enseñanza se alcanzaron y se utiliza para asignar calificaciones. Generalmente se define con tres criterios principales:

- Se determina a través de exámenes, pruebas, tareas o proyectos, que se usan para establecer que los estudiantes han aprendido, lo que deberían haber aprendido de acuerdo a las metas del curso y del plan de estudios.
- Se realizan al concluir periodos instruccionales específicos, y cubren el temario de ese periodo de tiempo.
- Se registran como puntuaciones o calificaciones que se incluyen en el registro académico del estudiante (que puede ser individual por parte del profesor o institucional por la universidad).

Lo que hace a una evaluación "sumativa" no es el diseño del examen o su contenido, sino la forma en que es utilizada para documentar cuánto han aprendido los estudiantes, para efec-

tos de calificación parcial, final, promoción, certificación, admisión o graduación (Miller et al, 2013). Este tipo de evaluación forma parte de la llamada "evaluación **del** aprendizaje", en contraste con la evaluación formativa (**para** el aprendizaje), que se describe en otro capítulo de esta obra. Una forma elegante e ingeniosa para entender los dos tipos de evaluación fue descrita por Paul Black: *"…cuando el cocinero prueba la sopa, esa es evaluación formativa. Cuando el cliente prueba la sopa, esa es evaluación sumativa"* (Black y Wiliam, 1998). En la Figura 1 se mencionan algunas de las diferencias entre estos dos tipos de evaluación.

Figura 1. Características de la evaluación sumativa y formativa



Desafortunadamente, con el paso del tiempo, se ha creado una separación artificial entre la evaluación sumativa y formativa, dando la impresión generalizada de que se trata de una dicotomía absoluta (Lau, 2016). A la evaluación sumativa se le ha etiquetado como meramente cuantitativa, que solo mide y no evalúa, que se centra solo en los números, que es punitiva y discriminatoria, usada con fines políticos, de ejercicio del poder o de control, y que es demasiado estandarizada para ayudar a los estudiantes en lo individual. Por el contrario, a la evaluación formativa para el aprendizaje se le describe como positiva, educativamente efectiva, que toma en cuenta los aspectos afectivos y emocionales de los estudiantes, y que ayuda a los educandos a salir adelante y aprender mejor, sin importar sus limitaciones personales y de contexto. Este debate creó una situación similar a la frase "*cuatro patas bueno, dos patas malo*", de George Orwell en "Rebelión en la Granja", que raya en lo absurdo al plantear límites arbitrarios que fragmentan innecesariamente estos complejos conceptos y generan divisiones epistemológicas que se convierten en irreconciliables (Lau, 2016).

Los dos tipos de evaluación deben visualizarse como un continuo, en el que las evaluaciones pueden tener componentes sumativos y formativos, dependiendo del uso que se dé

a los resultados (Bennett, 2015; Lau, 2016). Por ejemplo, un examen de ingreso a la universidad tiene un fuerte componente sumativo, aunque también puede usarse con fines diagnósticos para la institución y los docentes, así como formativos si se provee la información de los resultados a los estudiantes. Por otra parte, una reunión de retroalimentación con un estudiante durante el curso puede ser primordialmente formativa, aunque si la información obtenida en la entrevista de alguna forma cuenta para la calificación final, esta evaluación adquiere una dimensión sumativa. Ambos tipos de evaluación tienen virtudes y limitaciones, lo más razonable es utilizarlas con sensatez y sensibilidad de forma complementaria, de acuerdo a cada situación específica. Ejemplos de evaluación sumativa son los exámenes parciales o finales de curso, exámenes de certificación de profesionistas, exámenes de fin de carrera, exámenes de grado de maestría o doctorado, exámenes de admisión o de colocación. Estos exámenes tienen alta trascendencia para la vida del estudiantado, quienes a veces los perciben como obstáculos a sortear para alcanzar una meta, en lugar de oportunidades para progresar e identificar el nivel de su aprendizaje o competencia en un momento dado. Un tipo de exámenes sumativos que merece atención especial, son los llamados **"exámenes de alto impacto o de altas consecuencias"** ("*high-stakes testing*" en inglés), que ameritan una descripción aparte en este capítulo debido a sus implicaciones educativas (Sánchez y Delgado, 2017).

EXÁMENES DE ALTO IMPACTO (EAI)

Los exámenes de alto impacto o altas consecuencias tienen una larga historia en la educación superior a nivel internacional, y han contribuido en varias formas al desarrollo científico de la evaluación educativa como disciplina. A pesar de ello, el tema genera respuestas encontradas en varios sectores de la sociedad, docentes, estudiantes y profesionales de la educación, enfatizando con frecuencia sus potenciales efectos negativos. Todos los que hemos participado en procesos de ingreso y permanencia en educación superior, como profesores o estudiantes, hemos experimentado el impacto que pueden tener este tipo de exámenes. La aplicación de exámenes para ingreso a la universidad, para aprobar cursos, asignaturas y programas educativos, han sido parte integral del paisaje académico por mucho tiempo; el uso de instrumentos de medición que nos ubican en un nivel de desempeño específico, el suficiente para ser admitido en una institución, obtener una beca o tener acceso a estudios de posgrado e incentivos de diversos tipos, se han convertido en rutinas de evaluación que merecen reexaminarse a la luz de las investigaciones recientes y los efectos globales de la pandemia por COVID-19 (Cairns, 2021; Cizek, 2001; Tan et al, 2021).

Existen varias definiciones aceptadas por organizaciones internacionales sobre estos temas, a continuación, se describen algunas:

- **Examen o prueba** ("*test*" en inglés): "instrumento de evaluación que se emplea para identificar el nivel de dominio de los sustentantes sobre un constructo específico" (INEE, 2017). Los Estándares de AERA-APA-NCME lo definen como "recurso o procedimiento

en el que una muestra sistemática de una conducta del sustentante del examen en un dominio específico es obtenida y calificada utilizando un proceso estandarizado" (AERA, APA y NCME, 2014).

- **Exámenes de alto impacto (EAI) o de altas consecuencias** ("*high-stakes testing*" en inglés): "se indica cuando los resultados del instrumento tienen consecuencias importantes para las personas o las instituciones; por ejemplo, en los procesos de admisión o certificación" (INEE, 2017). En los Estándares de AERA-APA-NCME se definen como: "pruebas o exámenes cuyos resultados tienen consecuencias importantes y directas para los individuos, programas o instituciones involucrados en el examen" (AERA, APA y NCME, 2014). De acuerdo a la UNESCO: "evaluaciones con consecuencias importantes para los sustentantes, con base en su desempeño. Pasar el examen tiene beneficios importantes, como progresar al grado siguiente, obtención de un diploma o de una beca, ingresar al mercado laboral u obtener una licencia para practicar una profesión" (UNESCO, 2021).
- **Evaluaciones estandarizadas a gran escala** ("*large scale standardized testing*" en inglés): "son evaluaciones a nivel del sistema que proporcionan un resumen de los resultados de aprendizaje de un grupo de alumnos determinado, en un año académico determinado y en un número limitado de ámbitos. Suelen clasificarse como evaluaciones nacionales o transnacionales (regionales/ internacionales)" (UNESCO, 2021).

Como generalmente ocurre cuando se definen conceptos, establecer límites claros en constructos educativos y de ciencias sociales es problemático. Un examen puede ser de alto o bajo impacto dependiendo del contexto y de la percepción de la persona que presenta el examen, de quien establece inferencias de los resultados o de la misma institución que efectúa la evaluación. Un examen parcial de Estadística en una escuela de matemáticas podría ser de alto impacto si cuenta para la calificación final del estudiante y su beca depende del promedio académico, mientras que para un estudiante de filosofía que lleve Estadística como una asignatura opcional, el mismo examen puede ser de menor impacto.

IMPLICACIONES EDUCATIVAS DE LA EVALUACIÓN SUMATIVA Y LOS EAI

La discusión de la evaluación sumativa en educación es compleja y tiene varias aristas filosóficas, políticas, económicas, sociales. En particular los aspectos implícitos de la evaluación del aprendizaje y sus efectos sobre los procesos y fines de la evaluación, generan controversia por las convicciones o sesgos personales y grupales de quienes diseñan los instrumentos, de quienes los usan, de quienes son receptores de sus consecuencias, de la sociedad en su conjunto y de diferentes grupos de interés (Cizek, 2001; Márquez, 2014; Martínez Rizo, 2009; Nichols y Berliner, 2007; Sánchez Cerón et al, 2013). Cuando se discute la evaluación sumativa y los exámenes de alto impacto, se crea una retórica intensa que puede obstaculizar la discusión y dificultar el entendimiento.

La relativa escasez de conocimiento original empírico sobre los exámenes de alto impacto y sus efectos en el currículo, métodos de enseñanza y de estudio, dificulta una discusión

balanceada y de consenso, ya que la mayoría de las publicaciones sobre estos temas son artículos de opinión o textos que no han pasado el tamiz del arbitraje por pares, o son estudios con características muy locales y de contexto que hacen difícil su generalización (Sánchez y Delgado, 2017). La mayoría de los estudios sobre este tema están publicados en el litigioso contexto de Norteamérica, por lo que sigue siendo asignatura pendiente realizar investigación original sobre evaluación educativa en el contexto nacional y local, con perspectiva global.

De cualquier forma, existe una necesidad inescapable de tomar decisiones educativas. Por ejemplo, es prácticamente imposible que el 100% de la población acceda a la universidad en todos los países del mundo, y que realice la carrera profesional que cada individuo decida por preferencia personal, por lo que es poco probable vislumbrar un futuro cercano en el que no se realicen procesos de selección que incluyan exámenes sumativos de alto impacto. La sociedad requiere ser protegida de profesionistas y especialistas poco competentes, por lo que es difícil que desaparezcan totalmente los exámenes de certificación profesional, sobre todo en algunas disciplinas como la medicina y sus especialidades (Dauphinee, 2002; Brennan, 2006).

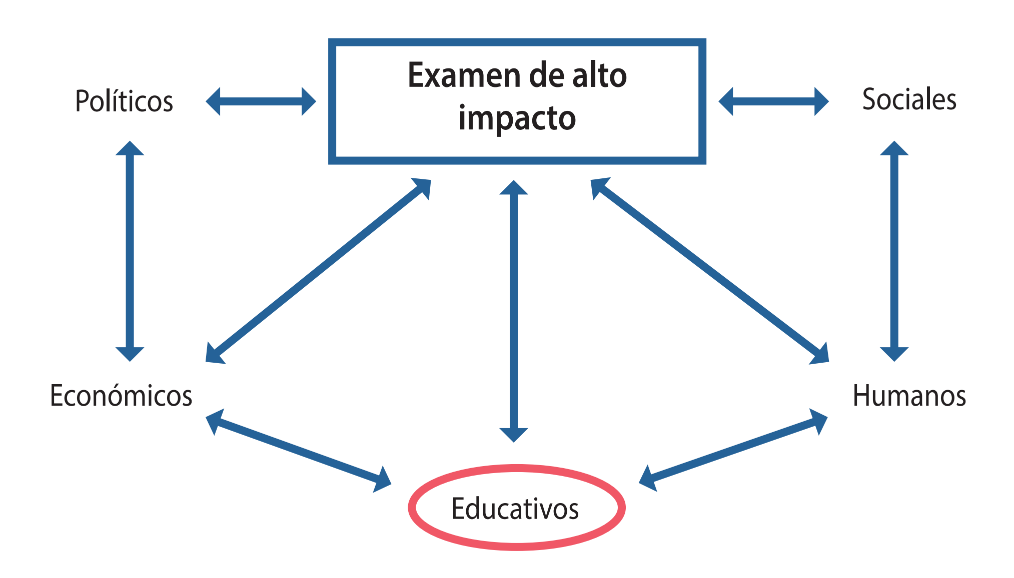
Uno de los principios de la evaluación educativa, es que no es recomendable tomar decisiones de alto impacto basadas en los resultados obtenidos con la aplicación de un solo instrumento. En este tipo de decisiones de gran trascendencia, es compromiso de las instituciones responsables realizar los exámenes de acuerdo a las buenas prácticas internacionales de evaluación educativa (AERA, APA y NCME, 2014; Brennan, 2006). Es inevitable que los métodos de evaluación utilizados en las universidades tengan efectos en estudiantes, profesores, el currículo formal y el oculto, entre otros (Brennan, 2006; Cizek, 2001; Haladyna y Downing, 2004). Los instrumentos de evaluación y el uso que se hace de ellos en las universidades son la declaración pública más importante de lo que "realmente cuenta" para la institución y la cultura de su cuerpo docente, y los estudiantes están alertas a ello.

Las implicaciones educativas y efectos de los exámenes de alto impacto son particularmente complejos, ya que generan señales a veces contradictorias que pueden distorsionar el proceso educativo y las prioridades del estudiantado y del profesorado. Como se muestra en la Figura 2, los efectos de los exámenes de alto impacto y sus consecuencias en las diferentes dimensiones del entorno educativo, funcionan como una red multidireccional de complejas interacciones en las esferas política, social, económica, humana y educativa.

Un ejemplo puede ser una reforma curricular de una escuela de medicina, en la que los métodos de evaluación propuestos en el plan de estudios deben estar alineados con el currículo, los métodos de enseñanza de los profesores, los resultados educativos esperados al final de la carrera y el contexto local y nacional de atención de la salud (Martone y Sireci, 2009). Cualquier cambio importante en los métodos de enseñanza de la escuela debe transitar por el complejo camino de implementación del cambio y adopción de innovaciones, procesos más sociales que técnicos, en los que se combinan aspectos políticos de las autoridades universitarias, el profesorado y los gremios disciplinarios, las necesidades de los estudiantes y las expectativas de la sociedad. Si predominan en el currículo los exámenes sumativos de

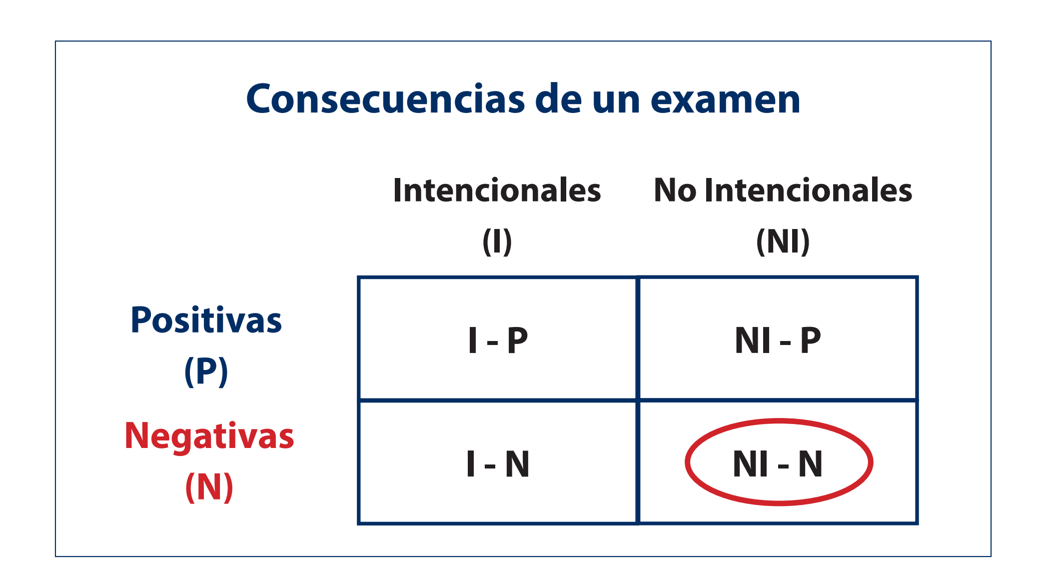
opción múltiple, aunque en el plan de estudios se declare que se enseñarán competencias genéricas, aspectos éticos, habilidades de comunicación, pensamiento crítico y creatividad, la motivación extrínseca de este tipo de exámenes y la sociología de su implementación puede influir dramáticamente en los métodos de estudio de los estudiantes, los contenidos enseñados en el aula y en el hospital, y la aparición de toda una industria de cursos para "preparar" a los estudiantes a obtener las puntuaciones más altas posibles en los exámenes. El currículo oculto se come de almuerzo al currículo formal cuando se trata de exámenes, es decir, la cultura vence a la estrategia (Newble y Jaeger, 1983; Madaus, 1988).

Figura 2. Esquema de las diferentes áreas en las que pueden tener consecuencias los exámenes de alto impacto (elaboración propia)



Para fines de este capítulo, describiremos algunos de los potenciales efectos educativos de los exámenes de alto impacto clasificándolos como positivos y negativos, en el entendido de que esta dicotomía es una simplificación de la realidad, y que estos efectos pueden convertirse en positivos o negativos dependiendo del contexto específico (Brennan, 2006; Sánchez y Delgado, 2017) (Figura 3).

Figura 3. Tipos de consecuencias de los exámenes de alto impacto, de acuerdo a su intencionalidad y direccionalidad (adaptado de Brennan RL, 2006)



A. Potenciales efectos positivos

- **Motivación para estudiar.** Los estudiantes identifican con eficacia cuáles evaluaciones "cuentan" para la calificación final, por lo que hacen su mejor esfuerzo en los exámenes parciales y finales (Martone y Sireci, 2009; Newble y Jaeger, 1983). Generalmente el estudiantado tiene una alta motivación intrínseca para aprender lo necesario para ser profesionistas exitosos, y los docentes deben ayudarlos haciendo explícitos los criterios de evaluación a utilizar en los cursos, para que los estudiantes tengan claridad sobre cómo serán evaluados. Diversos factores motivacionales extrínsecos e intrínsecos convergen en los estudiantes para dedicar esfuerzo a estudiar para los exámenes, lo que puede contribuir a mejorar el aprendizaje de los conceptos importantes del curso (Debray et al, 2003).
- **Estandarización de la evaluación.** Este es uno de los aspectos más controversiales de los exámenes de alto impacto. Los estándares de la AERA-APA-NCME afirman que es importante realizar los exámenes sumativos en condiciones estandarizadas, en ambientes consistentes, con reglas y especificaciones predefinidas detalladas, para que los contextos en que los sustentantes presentan el examen sean similares y las inferencias que se hagan de los resultados sean válidas (AERA, APA y NCME, 2014). En la edición más reciente de estos estándares se agregó el capítulo de justicia e imparcialidad ("*fairness*"), dándole el mismo nivel de importancia que a la validez y confiabilidad de los exámenes (AERA, APA y NCME, 2014). Varios autores han criticado la estandarización de la evaluación sumativa de alto impacto, en virtud de que los estudiantes constituyen

una población heterogénea y que los individuos son demasiado complejos para que su aprendizaje pueda ser realmente evaluado por este tipo de exámenes escritos de opción múltiple; uno de los argumentos es que estas pruebas evalúan primordialmente el conocimiento, excluyendo muchos otros aspectos importantes de las habilidades humanas para la vida y la práctica profesional (Márquez, 2014; Nichols y Berliner, 2007; Moreno-Olivos, 2010).

El debate persiste, aunque el peso de la evidencia empírica sugiere que los exámenes estandarizados, si son realizados profesionalmente, con uso apropiado y juicioso de los resultados, son una de las herramientas con mayor evidencia de validez y confiabilidad para identificar de manera justa el nivel de conocimiento, capacidad de entender conceptos y resolver problemas (AERA, APA y NCME, 2014; Brennan, 2006; Norcini et al, 2011; Sackett et al, 2008).

- **Mejora de la calidad educativa.** Si se siguen los lineamientos para realizar buenos exámenes, y se hace un esfuerzo por alinear la evaluación con el currículo y los métodos de enseñanza, es posible mejorar la calidad educativa (AERA, APA y NCME, 2014; Martínez Rizo, 2009; Yeh, 2005). La mejora de la calidad en el sistema educativo de un país depende de una red compleja de factores gubernamentales, sociales, económicos y personales de universidades, docentes y estudiantes, de los que los EAI son solo un componente. Los esfuerzos por mejorar la calidad educativa deben tener una perspectiva sistémica e identificar estrategias que, en el contexto local, incluyan a la evaluación del aprendizaje como un elemento fundamental.
- **Aprendizaje potenciado por exámenes** (capítulo 5 de este libro). Existe abundante investigación que documenta que realizar exámenes potencia el aprendizaje, más allá de su efecto motivacional directo para estudiar. A dicho concepto se le denomina "aprendizaje potenciado por exámenes" ("*Test-enhanced learning*"), es importante incorporarlo en nuestras estrategias evaluativas (Larsen et al, 2008).
- **Unificación de criterios.** El uso de EAI también puede contribuir a establecer consensos sobre diversos componentes de los procesos educativos, como los contenidos y metas educativas, la identificación de un currículo nuclear, el tipo de instrumentos a utilizar en evaluación, entre otros (Madaus, 1988; Martone et al, 2009). Este aspecto también es controversial, y puede generar rechazo en algunos docentes con el argumento de que se limita la libertad de cátedra.
- **Consecuencias positivas no intencionales.** Gregory Cizek, reconocido investigador en evaluación educativa, realizó una revisión de la literatura sobre las consecuencias positivas no intencionales de los exámenes de alto impacto, para poner en perspectiva la gran cantidad de artículos de opinión y anécdotas que enfatizan los aspectos negativos del tema (Cizek, 2001). Identificó los siguientes efectos positivos no intencionales de los exámenes de alto impacto:

    1) Desarrollo profesional. Las actividades de educación continua de los actores de la educación y evaluación han mejorado en calidad y efectividad.

2) Acomodación. Los exámenes de alto impacto han sido un catalizador para poner más atención a los estudiantes con necesidades especiales.
3) Conocimiento sobre evaluación. Los EAI han producido una mayor consciencia de la evaluación y su importancia.
4) Colección de información. Se han fortalecido los mecanismos de colección de datos e información generados en evaluación, y ha mejorado sustancialmente su calidad. Esto ofrece oportunidades para analizar información de forma longitudinal y transversal, con cruces de variables para informar la planeación de los procesos educativos y la toma de decisiones.
5) Uso de la información. Ligado al efecto anterior, el uso de la información generada en los EAI puede ser de utilidad para docentes y universidades.
6) Sistemas de rendición de cuentas. La rendición de cuentas en educación es un fenómeno que merece especial atención, en el que los EAI han jugado un papel relevante.
7) Familiaridad de los docentes con sus disciplinas. Las etapas para elaborar un examen de alto impacto implican varios pasos sucesivos e interdependientes, como la definición del perfil de referencia, tabla de especificaciones con resultados de aprendizaje definidos, participación de expertos en los temas a evaluar, entre otros. El diálogo que resulta de la interacción de docentes con expertos en evaluación durante el diseño de los exámenes, su análisis y retroalimentación, conduce a un incremento en el conocimiento sobre los temas actuales de su disciplina y cuáles son importantes para ser evaluados.
8) Calidad de los exámenes. La creciente utilización de exámenes de alto impacto ha llevado a mayor escrutinio en su diseño y análisis, lo que ha producido un incremento en la calidad técnica de los mismos. En palabras de Cizek: "por lo menos en términos de calidad técnica, el examen típico obligatorio de alto impacto que tome un estudiante será –por mucho– la mejor evaluación que el estudiante verá en todo el año" (Cizek, 2001).

B. Potenciales efectos negativos

La clasificación en efectos positivos y negativos de los exámenes de alto impacto no refleja necesariamente la compleja y dinámica realidad en la que un efecto puede ser benéfico o dañino dependiendo del contexto y otros factores mediadores. Sin embargo, creemos que es útil este esquema para tener en mente dichos efectos y reflexionar sobre su potencial impacto en los estudiantes.

- **Enseñando para la prueba.** Un efecto potencialmente negativo importante es el fenómeno denominado "enseñando para la prueba" ("*teaching to the test*") (Downing y Haladyna, 2004; Popham, 2001). El principal objetivo de las evaluaciones es obtener información que permita realizar inferencias sobre la adquisición de conocimientos y logros de las metas educativas definidas en el currículo, pero cuando docentes e instituciones enfatizan durante las actividades de enseñanza lo que vendrá en los exámenes,

entonces el proceso se distorsiona y puede llegar al grado de enseñar principalmente lo que vendrá en los exámenes. Incluso hay casos en que docentes y escuelas enseñan a sus estudiantes con preguntas que pueden venir en los exámenes, para mejorar las puntuaciones de grupos, escuelas y los mismos profesores. El aprendizaje se centra entonces en motivaciones extrínsecas, alterando el proceso educativo. El "enseñar para los exámenes" puede incluir varias actividades, desde las sutiles e inconscientes por parte del profesor, hasta las explícitas y dirigidas principalmente a subir las puntuaciones en los exámenes.

- **Cursos de preparación para exámenes.** Ante lo importante de las consecuencias de los EAI, han aparecido una gran cantidad de cursos, libros y plataformas digitales para mejorar las puntuaciones en los exámenes. Esto se ha convertido en un lucrativo negocio en México y el resto del mundo, aprovechando la necesidad de los aspirantes de aumentar sus posibilidades de aprobar y mejorar sus puntuaciones. En Estados Unidos, McGaghie y colaboradores realizaron una revisión sistemática sobre los cursos comerciales para preparar aspirantes para los exámenes de alto impacto en educación médica, en los que una sola empresa reportó ingresos por más de 250 millones de dólares (McGaghie et al, 2004). Encontraron que prácticamente no existe evidencia de su utilidad, los pocos estudios que muestran un efecto débil tienen una metodología de investigación deficiente, por lo que concluyen que no está demostrado que los cursos comerciales de este tipo tengan valor real.
- **Efectos en los currículos formal y oculto.** Existe controversia sobre el impacto de los exámenes de alta consecuencia en el currículo formal, vivido y oculto de las instituciones educativas (Au, 2007; Sackett et al, 2008). Las revisiones sistemáticas sobre el tema documentan que menos del 5% de la literatura publicada incluye datos empíricos, con metodología de investigación rigurosa, lo que hace difícil tener conclusiones contundentes. Tradicionalmente, se dice que las evaluaciones de alto impacto tienen influencia importante en el currículo, los métodos de enseñanza del profesorado y las estrategias de estudio de los estudiantes. Existe la percepción de que los graduados de las universidades tienen deficiencias en las habilidades necesarias para tener éxito en el mercado laboral, y que dedican demasiado tiempo y esfuerzo a contestar exámenes. Identificamos dos revisiones sistemáticas del tema, una de la Universidad de Michigan, EUA (Mehrens, 1998) y una meta-síntesis cualitativa de la Universidad de California (Au, 2007). Mehrens afirma que la totalidad de la evidencia no es clara y que depende del nivel del impacto o consecuencia del examen específico, y que no se ha logrado demostrar de manera contundente que los EAI influyan sustancialmente en el currículo, por lo menos en los trabajos de investigación cuantitativa, sin embargo, la comunidad docente persiste en la creencia de que los exámenes influyen en los planes de estudio, métodos de enseñanza y estrategias de estudio de los alumnos (Mehrens, 1998). En la meta-síntesis cualitativa de Au, se analizaron 49 estudios cualitativos sobre cómo los EAI afectan el currículo, los contenidos de conocimiento enseñados y las estrategias pedagógicas de los docentes, se encontró que el efecto principal de este tipo de exámenes es el estrechamiento del currículo, que se dirige a los contenidos examinados en las pruebas (Au, 2007). También encontró que las áreas de

conocimiento de los contenidos educativos se fragmentaban en piezas relacionadas con los exámenes, y que los docentes incrementaban el uso de estrategias pedagógicas centradas en el profesor, como la instrucción directa con conferencias y menor interactividad. Curiosamente, en una minoría de los estudios revisados por Au, algunos EAI tuvieron efectos positivos en las tres dimensiones citadas, con expansión del currículo, integración del conocimiento y estrategias de enseñanza centradas en el estudiante, por lo que el análisis sugiere que la naturaleza del control curricular inducido por los exámenes de alto impacto es dependiente de la estructura de los exámenes y del contexto (Au, 2007).

- **Inferencias inapropiadas de los resultados de los exámenes.** Uno de los efectos negativos más frecuentes de los exámenes de alto impacto es realizar inferencias de los resultados que no son congruentes con los objetivos iniciales del examen, por lo que dichas conclusiones tienen validez limitada (AERA, APA y NCME, 2014; Mendoza, 2015; Sánchez y Delgado, 2017). La elaboración e implementación de exámenes de alto impacto requiere gran inversión de recursos humanos y materiales, y el público usuario de la información con frecuencia no posee una cultura de evaluación suficiente para aplicar los conceptos de validez y confiabilidad. Con facilidad declaraciones breves y sensacionalistas en los medios de comunicación, generan malentendidos y distorsión de las conclusiones, limitaciones e implicaciones reales de los exámenes.

  El concepto moderno de validez es descrito en otro capítulo de este libro, como un modelo holístico en el que toda la validez es validez de constructo que se alimenta de diferentes fuentes, y que requiere de una cadena argumentativa para realizar inferencias apropiadas de los resultados; dicho concepto aún no ha permeado en la totalidad de la comunidad académica (AERA, APA y NCME, 2014; Downing y Haladyna, 2004). La comprensión de la validez es fundamental para entender las limitaciones de los resultados de los EAI, ya que extrapolar conclusiones y decisiones más allá de lo técnicamente correcto es inapropiado y puede ser peligroso y causar daño. Si un estudiante tiene un desempeño deficiente en un examen sumativo de alto impacto (como el examen profesional al graduarse de la carrera), eso no significa que sea una "mala persona", "incompetente", entre otros adjetivos que se asignan como etiquetas y que tienen un impacto emocional importante en los sustentantes. Una recomendación importante en evaluación educativa es: "Los desarrolladores del examen son los candidatos obvios para validar las afirmaciones que hacen sobre la interpretación de los resultados de un examen…" (Brennan, 2006), por lo que la responsabilidad de elaborar buenos instrumentos e informar a la sociedad sobre sus limitaciones recae en las instituciones responsables del examen.

EVALUACIÓN SUMATIVA Y EXÁMENES DE ALTO IMPACTO EN LA ERA PANDÉMICA

El impacto brutal de la pandemia por COVID-19 en la educación superior global ha sido particularmente fuerte en la evaluación sumativa y los exámenes de alto impacto, ya que muchos de ellos se tuvieron que suspender temporal o definitivamente en el año 2020, y, por otra parte, varios migraron a la modalidad de exámenes en línea (Cairns, 2021; Clark et al,

2020; Isbell y Kremmel, 2020; Luna-Bazaldua et al, 2020; UNESCO, 2020). Como ejemplo de lo profundo de la disrupción causada por la pandemia, uno de los exámenes más importantes en educación médica en Estados Unidos, el Paso 2 de Habilidades Clínicas del *United States Medical Licensing Examination* (Step 2 CS USMLE), examen clínico objetivo estructurado con estaciones estandarizadas y pacientes simulados para evaluar la competencia clínica, que es requisito para ejercer la profesión, así como para ingresar a una especialidad en EUA, fue temporalmente cancelado en un inicio y después suspendido definitivamente (https://www.usmle.org/work-relaunch-usmle-step-2-cs-discontinued), al no ser posible efectuarlo en las condiciones sanitarias vigentes.
Otro ejemplo espectacular ha sido la decisión de algunas de las universidades más famosas de los Estados Unidos como Harvard, Yale y la Universidad de California, de eliminar como requisito de admisión puntuaciones elevadas en los exámenes estandarizados como el SAT o el ACT (https://www.cnbc.com/2020/06/17/7-ivy-league-schools-will-not-require-sats-or-acts-next-year.html), como resultado de un prolongado debate, batallas legales y hallazgos de investigación que, aunado a los efectos de la pandemia en los estudiantes, parece anunciar un cambio profundo del paradigma del uso tradicional de los resultados de este tipo de EAI. Muchas organizaciones, que estaban en una relativa zona de comodidad en sus procesos de aplicación de exámenes estandarizados a gran escala, han tenido que repensar profundamente sus premisas, para iniciar una transformación de sus procesos y planes a corto y largo plazo, con el fin de enfrentar con éxito la nueva realidad trans y post-pandémica, así como aprovechar las herramientas de evaluación en línea que han madurado significativamente en los últimos años (Cairns, 2021; Clark et al, 2020; Isbell y Kremmel, 2020; Luna-Bazaldua et al, 2020; Tan et al, 2021; UNESCO, 2020).

Los retos de las brechas digitales y las dificultades para vigilar la administración de exámenes en línea en contextos menos controlados (como los exámenes en casa vigilados por herramientas digitales de video y algoritmos de inteligencia artificial) han generado un ambiente de rápidos avances y complejidades éticas y tecnológicas que no se han resuelto del todo; como se menciona en un documento de la UNESCO sobre el tema: "…los exámenes en línea solo se deben considerar cuando se hayan examinado exhaustivamente los temas de acceso igualitario a la infraestructura y la conectividad, la seguridad y los métodos de supervisión en línea, la transparencia, y las habilidades y brechas digitales de los estudiantes" (UNESCO, 2020). Luna-Bazaldúa y colaboradores recomiendan tener en mente los siguientes cinco puntos al mover los EAI a la modalidad en línea (Luna-Bazaldua et al, 2020):

1) No todo el estudiantado tiene acceso en casa a dispositivos y conexión a Internet adecuados para realizar un examen en línea, además de la situación personal de cada familia y su vivienda, que puede complicar responder adecuadamente este tipo de exámenes.
2) Pueden ocurrir problemas con el software, la electricidad y otras situaciones imprevistas. Se requiere tiempo y esfuerzo para probar la compatibilidad de los dispositivos con las plataformas de exámenes, así como apoyo técnico eficiente en tiempo real antes y durante el examen.

3) La vigilancia a distancia ("*remote proctoring*") en EAI con seres humanos en tiempo real, con programas de inteligencia artificial, o con grabaciones de video y revisión posterior de las mismas, tiene una gran cantidad de implicaciones éticas, legales, de invasión de la privacidad y manejo de datos personales, que no han sido resueltas en su totalidad (Coghlan et al., 2021).
4) No es suficiente convertir un examen previamente administrado en papel a un formato en línea, sin considerar todos los aspectos que implica la aplicación y análisis de un EAI (estudios piloto de los ítems, validez de contenido y de proceso de respuesta, comportamiento psicométrico de los reactivos, etc.) Debe tomarse en cuenta el formato en línea del examen durante todo el proceso de planeación e implementación.
5) Los EAI en línea deben seguir los principios de diseño universal, tarea nada sencilla en países como México, para permitir que el estudiantado con discapacidades tenga igualdad de oportunidades para demostrar sus habilidades.

La era moderna nos conduce a una oportunidad sin precedentes para mejorar nuestras estrategias de evaluación, y adecuar la evaluación sumativa y los exámenes de alto impacto a condiciones difíciles, dinámicas y complejas (Tan et al, 2021; Fuller et al, 2020).

CONCLUSIONES

Es imperativo desarrollar y aplicar instrumentos de evaluación siguiendo los principios fundamentales de evaluación educativa, utilizando las guías de cómo hacer exámenes escritos, basados en la mejor evidencia educativa disponible, para optimizar los efectos de los EAI en los individuos (estudiantado y profesorado), las instituciones, la sociedad y el desarrollo de los países. La posibilidad de afectar a los estudiantes por los resultados obtenidos en este tipo de evaluaciones es real y se ha convertido en un verdadero problema técnico, ético y de equidad. Un examen que no sigue los principios técnicos de diseño educativo puede ocasionar que aprueben estudiantes que no debieran pasar, y que estudiantes que merezcan aprobar no lo hagan, por lo que los beneficiarios de la profesionalización en evaluación educativa son múltiples: sustentantes, instituciones, docentes y la sociedad.

Debemos ampliar nuestros horizontes ante los retos de evaluaciones más auténticas y relacionadas al desempeño. Durning y colaboradores recomendaron recientemente implementar los conceptos de las ciencias de la complejidad y métodos no lineales en evaluación educativa, utilizando métodos innovadores para lograr la meta de evaluación basada en el trabajo y en el desempeño (Durning et al, 2015).

A continuación, anotamos algunas conclusiones:

- Los exámenes de alto impacto han contribuido al desarrollo de las ciencias de la evaluación.

- Estos exámenes tienen efectos positivos y negativos en los participantes de la educación superior, que son de naturaleza compleja y variable, dependiendo del contexto y de la naturaleza del examen.
- Las decisiones en los procesos de selección y de promoción tienen que tomarse, y es importante conocer las virtudes y limitaciones de los exámenes de alto impacto para incorporar la información generada en la toma sensata de estas ineludibles decisiones.
- La mayoría de las publicaciones sobre exámenes de alto impacto son opiniones, anécdotas o datos obtenidos sin metodología apropiada, por lo que es indispensable dedicar esfuerzos de investigación en esta temática, a nivel local y global.
- Los exámenes escritos tradicionales tienen utilidad limitada para explorar algunas de las habilidades necesarias para la vida (curiosidad, creatividad, empatía, compasión, resiliencia, entre otras) por lo que es necesario diseñar estrategias de enseñanza e instrumentos de evaluación apropiados para ello.
- Es necesario continuar innovando en el diseño, desarrollo y análisis de los exámenes de alto impacto, integrar la evaluación sumativa del aprendizaje con la evaluación formativa para el aprendizaje, para mejorar la calidad de la educación.

REFERENCIAS

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). Standards for educational and psychological testing. American Educational Research Association. https://www.testingstandards.net/open-access-files.html

Au W. (2007). High-Stakes. testing and curricular control: a qualitative metasynthesis. *Educational Researcher. 36*(5):258-267. https://journals.sagepub.com/doi/10.3102/0013189X07306523

Bennett, R. E. (2015). The Changing Nature of Educational Assessment. *Review of Research in Education*, *39*(1), 370–407. https://doi.org/10.3102/0091732X14554179

Black, P., Wiliam, D. (1998) Assessment and Classroom Learning, Assessment in Education: Principles, Policy & Practice, *5*(1), 7-74, DOI: 10.1080/0969595980050102

Brennan, R.L. (2006). Perspective on the Evolution and Future of Educational Measurement. En: Brennan, R.L., Ed. Educational Measurement. National Council on Measurement en Education and American Council on Education. 4th Ed. Westport, CT: Praeger Publishers, pág. 1-16.

Cairns, R. (2021). Exams tested by Covid-19: An opportunity to rethink standardized senior secondary examinations. *Prospects 51*(1-3), 331–345. https://doi.org/10.1007/s11125-020-09515-9

Cizek, G.J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice. 20*(4):19-27.

Clark, T.M., Christopher S. Callam, Noel M. Paul, Matthew W. Stoltzfus, and Daniel Turner Testing in the Time of COVID-19: A Sudden Transition to Unproctored Online Exams *J. Chem. Educ.* 2020, *97*(9), 3413–3417. https://doi.org/10.1021/acs.jchemed.0c00546

Coghlan, S., Miller, T. & Paterson, J. (2021). Good Proctor or “Big Brother”? Ethics of Online Exam Supervision Technologies. *Philos Technol*. 34, 1581–1606. https://doi.org/10.1007/s13347-021-00476-1

Dauphinee, W.D. (2002). Licensure and Certification. En: Norman, G.R., van der Vleuten, C.P.M., Newble, D.I. *International Handbook of Research in Medical Education*. Series: Springer International Handbooks of Education, *7*, 835-882.

Debray, E., Parson, G., Avila, S. (2003). Internal alignment and external pressure. En: Carnoy, M., Elmore, R., Siskin, L.S. (Eds.) *The new accountability: High schools and high-stakes testing*. New York: Routledge Falmer. pág. 55–85.

Downing, S.M., Haladyna, T.M. (2004). Validity threats: overcoming interference with proposed interpretations of assessment data. *Med Educ*. *38*(3):327-33. https://doi.org/10.1046/j.1365-2923.2004.01777.x

Durning, S.J., Lubarsky, S., Torre, D., Dory, V., Holmboe, E. (2015). Considering “nonlinearity” across the continuum in medical education assessment: supporting theory, practice, and future research directions. *J Contin Educ Health Prof*. *35*(3):232-243. https://doi.org/10.1002/chp.21298

Fuller, F., Joynes, V., Cooper, J., Boursicot, K., & Roberts, T. (2020) Could COVID-19 be our ‘There is no alternative’ (TINA) opportunity to enhance assessment? *Medical Teacher*, *42*(7), 781-786, DOI: 10.1080/0142159X.2020.1779206

Haladyna, T.M., Downing, S.M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*. *23*(1):17-27. https://doi.org/10.1111/j.1745-3992.2004.tb00149.x

Instituto Nacional para la Evaluación de la Educación. Criterios Técnicos para el Desarrollo y Uso de Instrumentos de Evaluación Educativa 2014-2015. INEE, México. 2017. https://www.inee.edu.mx/wp-content/uploads/2019/02/CRITERIOS_TECNICOS_PARA_EL_DESARROLLO_Y_USO_DE_INSTRUMENTOS_10_ABRIL_2014.pdf

Isbell, D. R., & Kremmel, B. (2020). Test Review: Current options in at-home language proficiency tests for making high-stakes decisions. *Language Testing*, *37*(4), 600–619. https://doi.org/10.1177/0265532220943483

Koretz, D.M., Linn, R.L., Dunbar, S.B., Shepard, L.A. (1991). The Effects of High-Stakes Testing on Achievement: Preliminary Findings About Generalization Across Tests. Presented at the annual meeting of the American Educational Research Association. En: Linn, L.R. *The Effects of High Stakes Testing, annual meeting of the American Educational Research Association and the National Council on Measurement in Education*, Chicago, IL; USA. https://dash.harvard.edu/bitstream/handle/1/10880553/The%20Effects%20of%20High-Stakes%20Testing%2023%20Dec%2002.pdf?sequence=1

Lau, A.M.S. (2016) ‘Formative good, summative bad?’ – A review of the dichotomy in assessment literature. *Journal of Further and Higher Education*, *40*(4), 509-525, DOI: 10.1080/0309877X.2014.984600

Larsen, D.P., Butler, A.C., Roediger, H.L. 3rd. (2008). Test-enhanced learning in medical education. *Med Educ*. *42*(10):959-66. https://doi.org/10.1111/j.1365-2923.2008.03124.x

Luna-Bazaldua, D., Liberman, J., and Levin, V. 2020. Moving high-stakes exams online: Five points to consider. Education for Global Development. https://blogs.worldbank.org/education/moving-high-stakes-exams-online-five-points-consider

Madaus, G.F. (1988). The influence of testing on the curriculum. En: Tanner, L.N. (Ed.), *Critical issues in curriculum: Eighty-seventh year-book of the national society for the study of education*. Chicago: University of Chicago Press. pág. 83–121.

Márquez Jiménez A. Las pruebas estandarizadas en entredicho. *Perfiles Educativos*. 2014; *36*(144):3-9. https://www.sciencedirect.com/sdfe/reader/pii/S0185269814706208/pdf

Martínez Rizo, F. (2009). Evaluación formativa en aula y evaluación a gran escala: hacia un sistema más equilibrado. *Revista Electrónica de Investigación Educativa*. *11*(2). http://redie.uabc.mx/redie/article/view/231

Martone, A., Sireci, S.G. (2009), Evaluating Alignment Between Curriculum, Assessment, and Instruction. *Review of Educational Research*. *79*(4):1332–1361. https://doi.org/10.3102/0034654309341375

McGaghie, W.C., Downing, S.M., Kubilius, R. (2004). What is the impact of commercial test preparation courses on medical examination performance? *Teach Learn Med. 16*(2):202-11. https://doi.org/10.1207/s15328015tlm1602_14

Mehrens, W. A. (1998). Consequences of Assessment: What is the Evidence? *Education Policy Analysis Archives*, *6*, 13. https://doi.org/10.14507/epaa.v6n13.1998

Mendoza Ramos, A. (2015). La validez en los exámenes de alto impacto: Un enfoque desde la lógica argumentativa. *Perfiles Educativos*. *37*(149):169-186. https://doi.org/10.22201/iisue.24486167e.2015.149.53132

Miller, M.D., Linn, R.L., Gronlund, N.E. (2013). *Measurement and Assessment in Teaching.* Pearson: USA. 11th Ed.

Moreno-Olivos, T. (2010). Lo bueno, lo malo y lo feo: las muchas caras de la evaluación. *Revista Iberoamericana de Educación Superior. I*(2):84-97. https://doi.org/10.22201/iisue.20072872e.2010.2.6

Newble, D.I., Jaeger, K. (1983). The effect of assessments and examinations on the learning of medical students. *Med Educ*. *17*(3):165-71. https://doi.org/10.1111/j.1365-2923.1983.tb00657.x

Nichols, S.L., Berliner, D.C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools.* Cambridge, MA: Harvard Education Press.

Norcini, J., Anderson, B., Bollela, V., Burch, V., Costa, M. J., Duvivier, R., Galbraith, R., Hays, R., Kent, A., Perrott, V., & Roberts, T. (2011). Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher*, *33*(3), 206–214. https://doi.org/10.3109/0142159X.2011.551559

Popham, W.J. (2001). Teaching to the Test? *Educational Leadership*, 58(6):16-20. http://www.ascd.org/publications/educational-leadership/mar01/vol58/num06/Teaching-to-the-Test%C2%A2.aspx

Sackett, P.R., Borneman, M.J., Connelly, B.S. (2008). High stakes testing in higher education and employment: appraising the evidence for validity and fairness. *Am Psychol*. *63*(4):215-27. https://doi.org/10.1037/0003-066X.63.4.215

Sánchez Cerón, M., del Sagrario Corte Cruz, F.M. (2013). Las evaluaciones estandarizadas: sus efectos en tres países latinoamericanos. *Revista Latinoamericana de Estudios Educativos (México)*. *43*(1):97-124. https://rlee.ibero.mx/index.php/rlee/article/view/285

Sánchez-Mendiola, M., Delgado-Maldonado, L. (2017). Exámenes de alto impacto: Implicaciones educativas. *Inv Ed Med* 6(21):52-62. http://dx.doi.org/10.1016/j.riem.2016.12.001

Tan C, Chua W, Vu CK, *et al*. (2021). High-stakes examinations during the COVID-19 pandemic: to proceed or not to proceed, that is the question. *Postgraduate Medical Journal*. *97*:427-431. https://pmj.bmj.com/content/97/1149/427

UNESCO (2021). Glossary. Paris: UNESCO. https://learningportal.iiep.unesco.org/es/glossary/e

UNESCO (2020). COVID-19. A glance of national coping strategies on high-stakes examinations and assessments. Paris: UNESCO. https://en.unesco.org/sites/default/files/unesco_review_of_high-stakes_exams_and_assessments_during_covid-19_en.pdf

Yeh, S.S. (2005). Limiting the unintended consequences of high-stakes testing. *Education Policy Analysis Archives*. *13*(43). http://scholarcommons.usf.edu/cgi/viewcontent.cgi?article=1577&context=coedu_pub

Capítulo 6

LA EVALUACIÓN DE LA DOCENCIA, PERSPECTIVAS DE UNA EXPERIENCIA INSTITUCIONAL

Laura Elena Rojo, Adrián Martínez González

"Los profesores afectan a la eternidad,
nadie puede decir dónde se termina su influencia."
Henry Brooks Adams

INTRODUCCIÓN

Este escrito expone las reflexiones de los autores sobre un tema esencial en los sistemas de educación superior alrededor del mundo que es el de la evaluación de la docencia; y describe, en sus líneas generales, la metodología desarrollada en la Dirección de Evaluación Educativa para valorar el desempeño de los docentes en distintas entidades académicas de la Universidad Nacional Autónoma de México (UNAM), a través de sus comunidades de estudiantes.

EL POR QUÉ Y PARA QUÉ DE LA EVALUACIÓN DE LA DOCENCIA

La tarea central de las instituciones de educación superior consiste preponderantemente en la formación de las nuevas generaciones de profesionales para los distintos campos del quehacer de la sociedad de nuestro tiempo. Esta tarea, expresada de manera rápida y sencilla entraña una extraordinaria responsabilidad social, porque debemos asumir que la misma tiene que realizarse con la máxima calidad para asegurar la salida de profesionales con sólidos conocimientos, habilidades, actitudes y valores para sostener sus realizaciones y encauzar transformaciones en sus respectivos escenarios de trabajo.

Una tarea educativa de este corte queda en manos de miles de docentes que día con día, en los espacios educativos –sean físicos o virtuales–, ponen en operación lo dispuesto en los planes y programas de estudio de las instituciones educativas; al hacerlo activan y conectan en la práctica sus conocimientos disciplinares y pedagógicos, su experiencia en el ejercicio de su profesión y en la docencia desde donde crean síntesis en las que se cristalizan interpretaciones pedagógicas de los conocimientos de uno o varios campos para lograr enseñarlos, tal y como

lo afirma Lee Shulman en sus estudios (Shulman, 2005). Facilitan el aprendizaje significativo a través de estrategias y técnicas que utiliza según las circunstancias y el nivel de desarrollo grupal de los estudiantes en el proceso educativo. La docencia, entonces es un ejercicio de máxima complejidad que interesa conocer de forma objetiva y sistemática.

A su vez, la docencia, como un ejercicio de responsabilidad social, tiene que ser revisada, examinada, sujetada a crítica, es decir, evaluada sistemática y periódicamente para disponer de evidencias sobre distintas aristas que la articulan y que resultan esenciales; como la formación del profesorado, su actualización, sus tareas institucionales, su desempeño en espacios educativos, su productividad pedagógica, entre otros aspectos.

En razón de ello, las instituciones educativas han impulsado sistemas, programas y estrategias de evaluación y a la luz de distintas lentes han tratado y tratan de apreciar de manera objetiva aspectos como los citados para diversas audiencias o grupos interesados, así como para sustentar decisiones de distinta índole (Ory, 2020).

Es larga la historia de los esfuerzos y discusiones en esta materia, podría decirse que es ya centenaria, ello ha dado lugar al desarrollo de numerosos marcos de referencia y modelos que definen, unos, los aspectos que deben ser valorados en torno al desempeño docente; en cambio, otros, definen el cómo pueden ser valorados. Al amparo de los primeros y los segundos las instituciones educativas han generado y puesto a prueba sistemas, programas y estrategias de evaluación de su planta académica (Sánchez Cabrero, 2021).

Los modelos que definen qué se debe evaluar de las y los docentes han estado inspirados en largos inventarios de atributos y conocimientos que estos deben dominar en su práctica educativa, algunos como el de Charlotte Danielson o el de Robert Marzano. (Danielson, 2011, Marzano, 2017) gozan de amplia popularidad y aceptación en los sectores educativos y, por lo tanto, su uso está extendido en distintas instituciones en varios países.

Otros autores y autoras y sus consecuentes discusiones se han concentrado en proponer, delinear y poner a prueba modelos de evaluación que se sustentan en la intervención de distintos actores y que, por lo tanto, generan evidencias y juicios acerca de la docencia de distinto carácter, rigor y peso. No tiene el mismo peso el juicio sobre el desempeño de la docencia de una o un académico prestigiado y reconocido, que el de una o un alumno en proceso de formación. Ello no significa que el segundo no sea de valía, simplemente la trayectoria y estatus de cada uno da lugar a un juicio distinto.

De entre estos modelos centrados en la participación de distintos actores de las comunidades educativas, los basados en el juicio de pares y los que se basan en la opinión de alumnado están totalmente arraigados en las instituciones educativas alrededor del mundo. El modelo basado en la opinión de alumnos, a pesar de las múltiples controversias de las que es objeto, prácticamente no hay institución universitaria que no lo considere a través del empleo de cuestionarios, mediante entrevistas a los alumnos, o a través de grupos focales.

El debate en las instituciones sobre qué y cómo evaluar el desempeño de sus docentes sigue estando a la orden del día, no está a juicio si evaluar o no, sino cómo hacerlo para poder disponer de información que permita decidir acciones de mejora, es aquí donde el debate continúa hasta nuestros días.

LA EVALUACIÓN DEL DESEMPEÑO DOCENTE EN LA UNAM. UN MOSAICO COMPLEJO

La Universidad Nacional Autónoma de México es una institución grande y compleja por la diversidad de los niveles educativos que opera, así como por la diversidad de terrenos disciplinarios en los que oferta sus planes y programas de estudio. Parte de su complejidad también se expresa en el número de escuelas, facultades, centros, institutos, y programas de posgrado que día con día ejercen una función educativa. En la actualidad, esta institución cuenta con tres bachilleratos: la Escuela Nacional Preparatoria con nueve planteles, el Colegio de Ciencias y Humanidades con cinco planteles y el bachillerato a distancia B@UNAM; para la formación de profesionales cuenta con veintinueve entidades académicas y, para formar posgraduados dispone de cuarenta y dos programas que, en su mayoría, ofrecen estudios tanto de maestría como de doctorado (UNAM, 2021). La complejidad institucional también tiene que ver con la autonomía académica de la que gozan las entidades que sostienen la función educativa, misma que les da la posibilidad de decidir acciones orientadas al cuidado de la calidad de sus esfuerzos. En esta autonomía académica, es justamente donde se sitúan las decisiones que las entidades toman para valorar el desempeño de su profesorado en los espacios educativos (aulas, laboratorios, unidades de atención a la salud, comunidades, entre otros). Otro tipo de evaluaciones, aquellas que tienen que ver con el ingreso y con las promociones, están reguladas por las disposiciones contenidas en el Estatuto de Personal Académico de la institución.

De esta manera, evaluar o no evaluar el desempeño del profesorado en los espacios educativos, ha sido una decisión que paulatinamente se ha ido tomando en la gran mayoría de las entidades académicas, tanto en las que ofertan estudios de bachillerato, como en las que sostienen los estudios profesionales y las que coordinan los estudios de posgrado; de tal suerte que es posible afirmar la existencia de múltiples y diversas experiencias de evaluación del desempeño docente en la UNAM, algunas añejas, que vienen desde la segunda mitad del siglo XX, otras más, actuales que comenzaron en la primera década del siglo XXI.

En la UNAM, los métodos y los procedimientos para el establecimiento de los sistemas o programas de evaluación del desempeño de las o los docentes en las aulas con la opinión del estudiantado pueden dividirse en dos familias.

La primera agruparía a las entidades cuyos programas son definidos, operados y actualizados con la participación solo de actores internos de las propias entidades, en este sentido, esta familia tiene una arista de naturaleza endogámica.

La segunda familia agrupa a las entidades académicas que han decidido impulsar sus programas con el apoyo de la Dirección de Evaluación Educativa (DEE) a través de un esquema de responsabilidad compartida. En esta familia, la DEE opera como instancia externa que ayuda a construir y operar el programa de evaluación del desempeño docente.

Cabe señalar que la UNAM es rica en experiencias en este campo y bien valdría la pena realizar un estudio comparativo de las mismas.

UNA METODOLOGÍA DE EVALUACIÓN EN PERFECCIONAMIENTO CONTINUO

Desde su fundación como Dirección General de Evaluación Educativa en febrero de 1997, esta dependencia de la Universidad Nacional Autónoma de México estableció una metodología para trabajar en conjunto con las entidades que le solicitaran apoyo para crear su *Programa de evaluación del desempeño docente* o solo desarrollar sus instrumentos de evaluación (Acuerdo de creación de la Secretaría de Planeación de la Universidad Nacional Autónoma de México, febrero 6 de 1997).

Las acciones en esta materia en la dependencia, desde su origen se fundamentan en seis ejes, a saber: 1. Filosofía de la evaluación; 2. Contextualización del servicio; 3. Desarrollo de instrumentos pertinentes bajo un esquema de corresponsabilidad; 4. Promoción de un ejercicio cíclico de precisión; 5. Sistematización de información útil para la entidad y los académicos evaluados; 6. Gestión compartida de procesos. Veamos cada uno de ellos.

El primero de los ejes, *una filosofía de la evaluación*, exigió determinar la orientación con la cual se daría un servicio tan delicado en la institución como lo es la creación de "*Programas para valorar el desempeño de los docentes mediante la opinión de su alumnado*". Considerando las implicaciones del modelo de evaluación basado en la opinión de las y los alumnos, así como el análisis de las distintas controversias que alrededor de él se plantean en las instituciones educativas. (Kulik, 2021), en la Dirección de Evaluación Educativa se acordó adoptar una filosofía de la evaluación formativa, esto es, impulsar programas de evaluación de naturaleza formativa cuyo propósito esencial fuese el de conocer aspectos selectos del desempeño del profesorado para poder apoyarlo con acciones de mejora en su actuación docente en las aulas. Asumir esta orientación fue y ha sido determinante para infundir confianza en todos los participantes, en particular, en los sujetos de la evaluación, las académicas y académicos de las entidades solicitantes del servicio (Goe, et. al., 2008).

El segundo eje, *la contextualización*, comprende caracterizar el servicio solicitado, condición que encierra considerar a cada entidad solicitante del servicio en varios aspectos: su historia y dinámica institucional; sus necesidades de evaluación; sus concepciones y prácticas alrededor de este proceso; sus incursiones en él; los atributos de sus planes de estudio, tales como modalidades de la docencia inscritas: docencia individual, a cargo de un único docente, docencia colectiva, a cargo de colectivos docentes o equipos de académicas y académicos, docencia teórica basada solo en el tratamiento de conceptos, marcos teóricos, conocimientos de metodologías de un campo disciplinario, docencia práctica, que pone a prueba lo aprendido en la teoría, o docencia teórico práctica que combina la teoría y su puesta a prueba en un mismo tiempo, entre otras, con el fin de situar correctamente el campo de acción de la evaluación, sopesar su complejidad y poder delinear condiciones que deben ser tomadas en cuenta para el diseño del programa de evaluación, así como de sus instrumentos específicos. Este eje es esencial para poder desterrar inclinaciones de los peticionarios para emplear instrumentos de otras instituciones o programas educativos, así como para argumentar la necesidad de contar con instrumentos basados en las características intrínsecas de la docencia que se ejerce en la entidad.

El tercero de los ejes, *desarrollo de instrumentos pertinentes*, significa trabajar con una entidad y con su comunidad académica para poder crear los instrumentos necesarios para respaldar la evaluación del desempeño de sus profesores y profesoras. El primero de los instrumentos es el perfil de la buena docencia (individual, colectiva, teórica, práctica o teórico-práctica) una definición institucional vital para muchos fines, no solo para la evaluación del desempeño docente. La sistematización de este instrumento puede hacerse a través de distintos procedimientos, uno de ellos, es mediante la consulta directa a las comunidades de docentes y de alumnas y alumnos de la entidad solicitante, a cuyos integrantes se les pide definir un conjunto de atributos sobre la buena docencia en el terreno disciplinario. Esta consulta puede hacerse en forma abierta a toda la planta académica y al estudiantado, o bien a una muestra o a través de grupos focales en los que se trabaja con muestras intencionales. El propósito es involucrar a la comunidad de la entidad en el proceso, recuperar su experiencia y expectativas y conformar, a partir de ello, el estándar ideal que tienen de una buena práctica en el ejercicio de la docencia en las aulas, con las características que estas pueden asumir, aulas físicas o aulas virtuales. Otro procedimiento, es el que se apoya en el análisis sistemático del plan de estudios, esto es del proyecto educativo rector en la entidad y de las prescripciones metodológicas contenidas en los programas de sus asignaturas, módulos o actividades académicas de donde se derivan conocimientos y acciones que debe dominar el cuerpo académico. Un tercer procedimiento implica la revisión de la literatura internacional y en la revisión de buenas prácticas institucionales. Cabe señalar que es posible, formular el perfil de la buena docencia combinando procedimientos, esto es, realizar la consulta, sumar los atributos de la docencia inscritos en los planes de estudios y, contrastar los conjuntos de atributos organizados en dominios de desempeño con una revisión de la literatura especializada del campo. Así, la Dirección de Evaluación Educativa a lo largo de los años ha incursionado en esta tarea con uno o varios de los procedimientos descritos, hoy por hoy, para todas las solicitudes que recibe en esta materia se integran perfiles anclados en los tres procedimientos: una consulta con la comunidad interesada, el análisis sistemático del plan de estudios y una revisión de la literatura especializada. El perfil propuesto de esta manera, antes de continuar con la generación de otros instrumentos, se somete a un revisión mediante el juicio de expertos en evaluación de la docencia, así una vez que el perfil está consolidado, se trabaja con él para determinar qué dominios y atributos pueden ser valorados mediante la opinión del alumnado. Identificados aquellos que son evaluables por este sector de la comunidad se formula la propuesta de cuestionario o en su caso de cuestionarios de evaluación, si en el caso en cuestión, la entidad declara querer evaluar cada tipo de docencia con especificidad.

Para el diseño de los cuestionarios la Dirección de Evaluación Educativa sigue un conjunto de reglas para especificar los reactivos que los formarán. Las describimos brevemente: a. que correspondan al dominio y atributos que le pertenecen al dominio; b. que se refieran solo a una acción realizada por el profesor o la profesora en el aula, observable por parte del alumnado; c. que estén expresados en un lenguaje comprensible por las alumnas y alumnos; d. que las especificaciones de las afirmaciones se expresen en términos afirmativos; e. que su longitud sea equivalente; f. que se correspondan con la escala de respuesta seleccionada.

Una vez que el diseño del contenido del cuestionario o los cuestionarios está completo, se formatean con identificadores institucionales e instrucciones específicas y se someten al juicio de especialistas en evaluación y así lograr la evidencia de validez en sus distintas fuentes (AERA 2014).

Como el trabajo de la DEE se hace bajo un esquema de corresponsabilidad, una vez que el perfil de la docencia y el cuestionario de evaluación, o los perfiles y cuestionarios, están completamente diseñados se elabora un documento de propuesta para la entidad solicitante a fin de que, en esta, sus autoridades unipersonales y colegiadas la sancionen y señalen precisiones que, a su juicio, deben introducirse en los instrumentos para evaluar a sus docentes. Conviene destacar que esta tarea en el proceso de creación de un Programa de evaluación del desempeño docente es crítica, puesto que los dos instrumentos a los que nos hemos referido –el perfil de la buena docencia y el cuestionario de evaluación– constituyen el timón del ejercicio evaluativo.

El cuarto eje, *promoción de un ejercicio cíclico de precisión*, tiene que ver con visualizar la evaluación del desempeño docente como un ejercicio de ciclos, periódico, con criterios y reglas claros. Esto nos conduce a que la creación de cualquier *Programa de evaluación del desempeño docente* que apoya la Dirección de Evaluación Educativa se asuma en la entidad solicitante como un ejercicio sistemático y cíclico, nunca como un ejercicio coyuntural para dar respuesta a peticiones o a presiones internas o externas que no consideran el proceso evaluativo como un componente vital, de naturaleza continua, en la vida académica de la institución.

En este sentido, el diseño de un programa de evaluación se hace con perspectiva de tiempo, con el propósito de que una entidad acumule evidencias del desempeño de su profesorado, que con fundamento en aplicaciones sucesivas del instrumento se puedan establecer trayectorias de desempeño docente útiles para fortalecer las decisiones de mejora de la docencia en la institución. Cuando en este eje, también se incluye el aspecto de la precisión significa que las evaluaciones deben planearse para poder realizarse en los tiempos en que las y los alumnos estén libres de emociones y presiones por sus evaluaciones, o libres también de las cargas de trabajo que suponen el final de un ciclo de estudios. La precisión también tiene que ver con la acción puntual de un conjunto de actores que desatan tareas vinculadas con la evaluación que más adelante describimos.

El quinto eje, *la sistematización de información útil,* en la metodología desarrollada por la Dirección de Evaluación Educativa de la UNAM se asocia con la producción de resultados que deben ser devueltos a las entidades que solicitan el servicio, así como a las y los docentes evaluados por su grupo de estudiantes.

En este componente se asume que la práctica del ejercicio evaluativo tiene que cerrarse con la devolución de información específica y comprensible, apegada a los objetivos de la evaluación, sistematizada en formatos que sean tanto claros y precisos, de utilidad en varias escalas: para la entidad demandante del servicio, para sus cuerpos colegiados, para sus áreas académicas y, sin lugar a duda para sus docentes.

La DEE a lo largo del tiempo ha perfeccionado dos instrumentos para presentar resultados, el informe de evaluación de corte institucional, y el informe personal para las y los profesores evaluados.

El primero de ellos –el informe institucional– da cuenta y reseña, por cada ciclo de evaluación, el marco de referencia institucional, para contextualizar el ejercicio de evaluación; sus objetivos; los instrumentos empleados para la evaluación y en su caso, las reformas introducidas en estos; la población o, si se hubiese decidido, la muestra de docentes determinada por la entidad solicitante; el proceso de aplicación: número de profesores evaluados y número de alumnas y alumnos participantes; el análisis de los datos, los resultados globales por la entidad, por sus áreas, por tipo de asignatura, módulo o por modalidad de docencia. Este informe permite visualizar los resultados globales y ofrece la posibilidad de identificar el tipo de decisiones y acciones institucionales que conviene realizar para apoyar la mejora de la actuación del profesorado en las aulas. En muchos casos, es el informe que puede compartirse con otras instancias institucionales, incluso, con agentes externos a la institución, como pueden ser los organismos acreditadores.

Por su parte, el informe personal es el informe que se prepara para cada profesora o profesor evaluado. En este se detallan los resultados obtenidos en el ciclo escolar y se incluye una orientación para interpretar correctamente los resultados con el fin de que cada docente reflexione a partir de ellos sobre su desempeño en el aula.

El sexto y último eje, *la gestión compartida de procesos*, comprende la logística que necesariamente se tiene que considerar en todo proceso de evaluación y que encierra las tareas de planeación, desarrollo y revisión de toda la evaluación. En este sentido, a través de este componente y en cada ciclo de evaluación la DEE, junto con cada entidad: planea y programa el proceso; aseguran sus insumos, disponen, revisan y actualizan los sistemas de evaluación, cuando se hace en línea; se difunde el proceso en la entidad; se realizan invitaciones y recordatorios a quienes asumen el papel de evaluadores, en este caso, los alumnos que emiten su opinión basada en su experiencia directa con sus docentes; se acuerda el monitoreo de los sistemas de línea y de los procesos; se registran y resuelven incidentes, entre otras muchas acciones necesarias. Valga la pena señalar que un Programa de Evaluación del Desempeño Docente exige movimientos múltiples en secuencias de tiempo precisas, se pone en juego, la obtención de evidencias para soportar en ellas decisiones de mejora, esenciales en la dinámica educativa de la universidad, por esta razón la gestión compartida de los procesos, ciclo con ciclo, es también un componente que merece gran atención.

Hasta aquí la descripción de la metodología desarrollada e instrumentada por la DEE sobre Evaluación de la Docencia, a su perfeccionamiento han contribuido la voz y las propuestas de los grupos con los que se tienen acuerdos para apoyar la operación de sus programas de evaluación del desempeño docente.

Como es de esperarse una metodología para ser efectiva tiene que revisarse continuamente y debe perfeccionarse para beneficio de todas las partes.

LOGROS Y RETOS

Entre 2000 y 2021, la hoy Dirección de Evaluación Educativa ha colaborado con 72% de las facultades y escuelas que ofertan programas de licenciatura. De estas, en 43% de los casos ha apoyado el nacimiento de sus programas de evaluación del desempeño docente hasta lograr su estabilidad, o bien, ha emitido dictamen y recomendaciones sobre sus procesos e instrumentos específicos. En el resto de los casos, 57% mantiene acuerdos activos de asesoría y apoyo técnico en la operación completa y continua de sus programas de evaluación del desempeño docente. En algunos casos, se han introducido actualizaciones en los instrumentos en concordancia con la modificación de los planes de estudio. En cuanto a los programas de posgrado de la institución, la DEE mantiene acuerdo vigente con 55%, lo cual implica asesoría y apoyo técnico en la operación de sus programas de evaluación del desempeño docente.

Sin lugar a duda, un logro de la dependencia ha sido la configuración de una metodología de trabajo a partir de un esfuerzo colaborativo de los integrantes de la DEE, condición que ha permitido reforzarla y perfeccionarla con las aportaciones de los especialistas que forman parte su equipo de trabajo.

Un logro extendido en todas las entidades que han sido y hoy son apoyadas por esta dependencia ha sido la institucionalización progresiva del proceso de evaluación del desempeño docente considerado, en la actualidad, como un ejercicio relevante en la vida académica de cada entidad, así como su aceptación por parte de las comunidades de docentes, alumnas y alumnos.

Otro logro, indudable, ha sido el sostenimiento y búsqueda de perfeccionamiento continuo de las interacciones con las entidades, aspecto de suma fragilidad en la vida de una institución tan grande y tan compleja como lo es la UNAM.

En materia de retos, estos han convivido con la trayectoria de la antes Dirección General de Evaluación Educativa, actual Dirección de Evaluación Educativa desde su origen: uno de ellos, que merece ser considerado, es el ofrecer un servicio útil para las entidades académicas, con calidad y profesionalismo, arropado continuamente con la ética que cualquier ejercicio de evaluación demanda.

CONCLUSIONES Y REFLEXIONES FINALES

Los métodos de evaluación docente han evolucionado, la tendencia es contar con modelos válidos, precisos, prácticos, aceptables y congruentes con la filosofía institucional. Es relevante hacer un mayor reconocimiento a las y los docentes, que las instituciones valoren esta función y promueva el cambio hacia una cultura de la evaluación docente como un proceso formativo, constructivo y de perfeccionamiento del quehacer educativo, lo que implica trabajar en la construcción de instrumentos de evaluación objetivos, con evidencia de validez y confiabilidad. Hay que explorar nuevos horizontes de la evaluación docente, hacia la construcción de buenas prácticas de la evaluación del desempeño del profesorado.

La cultura de evaluación implica que se fortalezcan los sistemas de valores institucionales a fin de contrarrestar las actitudes negativas hacia la evaluación y crear un clima de

confianza y colaboración a lo largo de todo el proceso. En una cultura de evaluación donde se tenga como propósito profesionalizar la docencia y orientarla hacia la docencia basada en evidencias, es decir, utilizar mecanismos de enseñanza-aprendizaje que hayan sido útiles para los estudiantes, y que hayan sido diseñados y estudiados con rigor metodológico apropiado, para ello, la evaluación resulta un mecanismo necesario e imprescindible para la mejora, innovación y autorregulación de la calidad del desempeño docente y del propio sistema educativo.

REFERENCIAS

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). The standards for educational and psychological testing. Washington, D.C.: American Educational Research Association.

Danielson; Ch. (2011). *The framework for teacher evaluation.* The Danielson Group. Princeton, NJ.

Goe, Bell y Little. (2008). Approaches to evaluating teacher effectiveness: A research synthesis. https://gtlcenter.org/sites/default/files/docs/EvaluatingTeachEffectiveness.pdf Octubre, 2021.

Kulik, James. (2021). Student ratings: Validity, utility and controversy. En: *New Directions for Institutional Research, 109)*

Marzano, R. (2017). *The Marzano Teacher Evaluation Model and the Marzano Focused Teacher Evaluation Model,* https://www.marzanocenter.com/wp-content/uploads/sites/4/2019/04/FTEM_Updated_Michigan_08312017.pdf. Octubre, 2021.

Ory, J. (2000). Teaching evaluation: Past, present and future. En: *New Directions for Teaching and Learning, 83*. https://onlinelibrary.wiley.com/doi/epdf/10.1002/tl.8302. Octubre .2021.

Sánchez Cabrero, et. al. (2021). Models on teaching effectiveness in current sicientific literature. En: *Education Sciences, 11*, 409. https://files.eric.ed.gov/fulltext/EJ1307408.pdf Octubre 2021.

Shulman, Lee. (2005). Conocimiento y enseñanza: fundamentos de la nueva reforma. En: *Revista de curriculum y formación del profesorado, 9*, 2. https://www.ugr.es/~recfpro/rev92ART1.pdf. Octubre10, 2021.

UNAM. DGP. (2021) Agenda Estadística 2021. https://www.planeacion.unam.mx/Agenda/2021/disco/index.html. Octubre 2021.

Capítulo 7

APRENDIZAJE POTENCIADO POR LA EVALUACIÓN: UNA PRÁCTICA PARA PROMOVER EL APRENDIZAJE DEL ESTUDIANTADO

Melchor Sánchez Mendiola, Sandra Paola Rodríguez Castillo,
Nancy Fabiola Pérez Herrera

"La clave es que todos aprendamos a aprender de manera más eficaz."
Robert Bjork

INTRODUCCIÓN

¿Por qué hacer exámenes?, ¿para qué nos sirven?, ¿cómo evaluar lo que ha aprendido un estudiante? Estas pueden ser algunas de las preguntas que los profesores se hacen durante su práctica. Habitualmente, se piensa que las pruebas únicamente permiten evaluar el conocimiento de los alumnos durante o hacia el final de la revisión de un tema, unidad o curso, esto es, se utilizan comúnmente con fines formativos o sumativos y, de esta manera, se verifica el progreso hacia las metas de aprendizaje. No obstante, pueden servir para otros propósitos educativos, al brindar la oportunidad de enriquecer la forma en que se aprende; ya lo señala la investigación en psicología cognitiva, a medida que las evaluaciones se aproximan, de manera indirecta promueven conductas de estudio. Este efecto también se puede observar una vez que los alumnos reciben los resultados alcanzados y dirigen esfuerzos adicionales en áreas de bajo rendimiento (Green et al., 2018).

El impacto de la evaluación en la enseñanza y el aprendizaje no es nuevo, se habla de ella como un punto clave del proceso, dado que a partir de sus resultados se toman decisiones pedagógicas de gran relevancia (Widiastuti et al., 2020). A diferencia de las evaluaciones únicamente de tipo sumativo, la evaluación formativa se realiza de forma paralela y articulada a la enseñanza y aprendizaje, ofreciendo oportunidades tanto para los estudiantes como para los profesores, ya que no solo pretende la adaptación del aprendizaje, sino también el uso de diversas estrategias de enseñanza orientadas a promover mayores niveles de logro en los aprendizajes. Por tanto, toma fuerza al concebirse como una herramienta que permite recolectar información para identificar el progreso en el aprendizaje de los estudiantes, no solo en contextos presenciales, sino también en educación a distancia.

Este capítulo pretende subrayar cómo la evaluación, principalmente a través de exámenes o pruebas, puede mejorar el aprendizaje de los estudiantes, especialmente porque favorece tanto la recuperación como la retención de información, lo cual se conoce como aprendizaje potenciado por las pruebas (*test-enhanced learning*, en inglés) (Adesope et al., 2017; Roediger et al., 2011). Particularmente, se aludirá a exámenes o ejercicios que se realizan en los espacios educativos como parte del curso que se esté desarrollando, con la finalidad de que el profesorado lo pueda implementar cuando sea pertinente. En el capítulo 1 de este libro abordamos los diversos matices de la evaluación desde los puntos de vista del, para y como aprendizaje, enfatizando la importancia de considerar a la evaluación como parte integral del proceso de enseñanza y aprendizaje, en lugar de verla como algo separado y fragmentado.

La información que a continuación se presenta está organizada en cinco apartados. En el primero, se explica qué es el aprendizaje potenciado por las pruebas; en el segundo, se comenta lo que dice la investigación acerca de este fenómeno; en el tercero se establecen los beneficios de su aplicación; en el cuarto se proponen algunas recomendaciones para que los docentes puedan implementarlo en los espacios educativos. Por último, se plantean conclusiones y reflexiones finales.

Se espera que este material sirva como punto de partida para que el profesorado conozca las posibilidades que ofrece la evaluación como una práctica para potenciar el aprendizaje de los estudiantes.

¿QUÉ ES EL APRENDIZAJE POTENCIADO POR LA EVALUACIÓN?

La atención en las pruebas usualmente se centra en su propósito para valorar el nivel de logro, para emitir una calificación o para tomar decisiones de alto impacto, es decir, como herramienta de evaluación sumativa, pero esta visión opaca otras funciones importantes, por ejemplo, poco se habla de su efecto facilitador o potenciador del aprendizaje (Adesope et al., 2017; Roediger et al., 2011).

El creciente y necesario interés por comprobar la utilidad de las pruebas desde un marco de evaluación formativa, más allá de concebirse como un instrumento para la calificación, ha llevado a diversos autores a comprobar que su efecto no solo favorece, sino que potencia el aprendizaje de los estudiantes. Así, de forma repetida se escucha con mayor fuerza la frase en la que se expresa que *"las pruebas son una herramienta poderosa para el aprendizaje eficaz en entornos educativos"* (Choi y Lee, 2020). Lo anterior, cuenta con suficiente sustento empírico, desde diversos y exhaustivos análisis académicos.

Como parte de una evaluación formativa, un cuestionario de bajo impacto o *quiz* puede dar idea del avance o las dificultades presentes en el proceso educativo y, a manera de diagnóstico, un examen permitirá determinar los conocimientos y habilidades preexistentes en los educandos. Estudios indican que examinar de manera recurrente a través de pruebas o exámenes conlleva a la recuperación de información, conduce a una mayor retención e incluso, mejora la capacidad de recordar hechos relacionados. Este hallazgo es conocido

en psicología cognitiva como aprendizaje potenciado por las pruebas o efecto de las pruebas (*test-enhanced learning* o *testing effect* en inglés) (Adesope et al., 2017; Brame y Biel, 2015; Green et al., 2018; Roediger y Karpicke, 2006).

Aun cuando las pruebas de práctica o recuperación pueden considerarse una dificultad deseable, esto es, que pueden obstaculizar el aprendizaje a corto plazo, pero mejorarlo a largo plazo (Bjork, 1994 citado en Greving y Richter, 2018), es importante reflexionar por qué se está evaluando, qué información se quiere obtener y cómo se pueden utilizar los exámenes durante las clases como una estrategia para promover el aprendizaje. Por ejemplo, si los alumnos son evaluados cada semana, probablemente estudiarán más durante el semestre que si solo se aplicara un examen parcial y/o final, lo cual tendrá un efecto sobre las actividades de estudio.

Por lo tanto, para los fines de este capítulo, nos enfocaremos en las pruebas que se realizan como parte del curso o en las autoevaluaciones que los estudiantes pueden usar como estrategia de estudio.

¿CUÁLES SON LOS HALLAZGOS DE LA INVESTIGACIÓN?

Si bien diversos estudios en el tema se han realizado en entornos controlados de laboratorio, los investigadores buscan generalizar los resultados al espacio educativo, para lo cual han explorado si existe un efecto de las pruebas de práctica y también en qué condiciones estas pruebas mejoran o inhiben el aprendizaje. En esta sección se presentan factores como el formato o tipo de reactivos, la retroalimentación, la frecuencia con que se evalúa, la ansiedad ante los exámenes y su relación con el aprendizaje potenciado por la evaluación (Tabla 1).

Tabla 1. Factores que influyen en los efectos de las pruebas en el aprendizaje de estudiantes

Factores	Hallazgos de investigación
Formato de reactivos	Pruebas de opción múltiple son más útiles para memorización y retención de datos, pruebas de respuesta corta requieren pensamiento de orden superior y son útiles para contenido conceptual y abstracto.
Retroalimentación	Puede actuar como evaluación formativa, se confirman las respuestas correctas y se da información sobre preguntas que se respondieron incorrectamente.
Frecuencia de la evaluación	Práctica distribuida o espaciada (efecto espaciador), es más efectiva que la práctica masiva o en una sola sesión.
Ansiedad ante exámenes	Cuestionarios de bajo riesgo disminuyen la ansiedad ante los exámenes, como preparación para exámenes finales o sumativos.

Formato de reactivos

El tipo de reactivo que se utilice en las pruebas o exámenes puede influir en el resultado final. En consecuencia, los ítems de respuesta corta, opción múltiple y falso o verdadero resultan útiles en distintos tipos de contenido (Clair y Jensen, 2020). Por su parte, Little y Bjork (2015) y Collins et al. (2018) indican que preguntas de opción múltiple con buenos distractores que además se acompañan de comentarios, resultan una herramienta efectiva para el aprendizaje.

McDaniel et al. (2007) en uno de los primeros estudios dentro del contexto escolar donde los estudiantes resolvieron, semanalmente, cuestionarios de respuesta breve u opción múltiple, o volvieron a estudiar el contenido aprendido, en cualquier caso, acompañado por retroalimentación, encontraron que en una prueba criterial posterior, los ítems de respuesta breve produjeron un efecto de prueba más pronunciado que los de opción múltiple. Lo anterior coincide con Greving y Richter (2018), quienes sugieren que las pruebas de respuesta breve pueden beneficiar el aprendizaje en contextos de educación superior, en comparación con volver a revisar los puntos clave de la clase o los exámenes de opción múltiple.

En otros estudios se ha identificado que las pruebas de opción múltiple son más útiles para la memorización y la retención de hechos, mientras que las pruebas de respuesta corta requieren pensamiento de orden superior y son útiles para contenido conceptual y abstracto. Cuando se utilizó una prueba con reactivos de formato mixto, se halló que el efecto en el aprendizaje fue todavía mejor, lo cual se puede deber a que los estudiantes, ante diferentes formatos, llevan a cabo distintos procesos cognitivos (Adesope et al., 2017).

Ahora bien, ¿es importante que el formato de la prueba de práctica y la final sean idénticos? Esta pregunta se relaciona con el concepto de procesamiento apropiado para la transferencia (*transfer-appropriate processing o TAP* en inglés), según el cual, el desempeño en cualquier tarea dada será más alto si las características del procedimiento de aprendizaje son similares a las de la evaluación. De hecho, la razón por la que un examen de práctica es una herramienta útil para la preparación hacia los exámenes finales puede deberse a que imita los procesos mentales que los alumnos deben realizar en el examen real, así mismo, entre más profundo es el procesamiento cognitivo solicitado en la prueba de práctica, el desempeño de los estudiantes en futuras pruebas es mejor (Adesope et al., 2017; Collins et al., 2018).

Tabla 2. Algunas recomendaciones y recursos para el uso del aprendizaje potenciado por pruebas

Recomendación	Recursos
Preguntar durante la clase	https://citl.illinois.edu/citl-101/teaching-learning/resources/teaching-strategies/questioning-strategies https://veo.co.uk/effective-classroom-questioning-tips/
Uso de quizzes o cuestionarios frecuentes	https://www.insidehighered.com/advice/2020/11/04/how-make-online-quizzes-more-effective-opinion https://quizizz.com
Estrategias de recuperación en espacios educativos	https://eab.com/insights/blogs/district-leadership/strategies-to-expedite-learning-recovery/ https://www.heacademy.ac.uk/system/files/resources/Small_group_tea-ching_1.pdf
Usar puntos resumen	https://lincs.ed.gov/state-resources/federal-initiatives/teal/guide/teachsumm https://www.facultyfocus.com/articles/effective-teaching-strategies/let-students-summarize-the-previous-lesson/
Utilizar cuadrículas de desafío o recuperación	https://www.learningandthebrain.com/blog/retrieval-grids-the-good-the-bad-and-the-potential-solutions/ https://lovetoteach87.com/2018/01/12/retrieval-practice-challenge-grids-for-the-classroom/
Usar tarjetas de memoria o *flashcards*	https://www.learningscientists.org/blog/2016/2/20-1 https://usm.maine.edu/agile/using-flashcards
Aplicar la práctica distribuida	Study strategies: retrieval practice https://www.youtube.com/watch?v=Pjrqc6UMDKM http://pdf.retrievalpractice.org/RetrievalPracticeGuide.pdf
Informar sobre el aprendizaje potenciado por la evaluación	Test-Enhanced Learning Made Easy https://youtu.be/40ibPH17z4s Test enhanced learning https://sites.google.com/a/uwlax.edu/exploring-how-students-learn/test-enhanced-learning

Retroalimentación

Se ha realizado un trabajo considerable para examinar el papel de la retroalimentación sobre el aprendizaje potenciado por la evaluación, fenómeno que es efectivo en sí mismo, pero acompañado de ella mejora aún más los resultados (Lavigne y Risko, 2018, citados en Clair y Jensen, 2020). La evidencia empírica de los estudios de laboratorio parece respaldar esta práctica, Butler y Roediger (2008) diseñaron un experimento en el que estudiantes universitarios estudiaron 12 pasajes históricos seguidos por un examen de opción múltiple. Se dividieron en tres grupos, sin retroalimentación, aquellos donde fue inmediata (después de cada pregunta) y a quienes se entregó diferida (después de completar la prueba de 42 ítems). Una semana más tarde, los alumnos regresaron para una prueba completa de recuperación con pistas. Si bien completar las preguntas de opción múltiple después de leer los pasajes mejoró el rendimiento en la prueba final, recibir retroalimentación proporcionó un beneficio adicional. Cabe señalar que, la entrega diferida *versus* la inmediata resultó en un mejor rendimiento final, aunque ambas condiciones mostraron beneficios sobre su ausencia.

Aunque el meta-análisis realizado por Adesope et al. (2017), provee evidencia para señalar que el efecto de la prueba para mejorar el aprendizaje se observa independientemente de si se proporciona retroalimentación, el ofrecerla puede actuar como una evaluación formativa, en la medida que se confirman las respuestas correctas y se brinda información sobre las preguntas que se respondieron incorrectamente, aumentando así los efectos positivos de los exámenes. Un importante efecto en el aprendizaje al realizar pruebas o exámenes, que ha sido consistente en las investigaciones, es que esta estrategia es superior a "volver a estudiar", una de las maneras más tradicionales en las que el estudiantado se prepara para los exámenes sumativos. Es decir, ayuda más para el aprendizaje profundo y duradero el realizar exámenes y pruebas, que leer una y otra vez el material, lo que tiene implicaciones importantes para el uso del tiempo de las y los estudiantes.

Frecuencia de la práctica y la evaluación

¿Cuántas veces los profesores escuchan a los alumnos decir que estudiaron un día antes del examen?, ¿los alumnos estarán acostumbrados a estudiar al menos una hora por día o varias horas antes de una prueba?, ¿qué impacto tienen el tiempo y frecuencia del estudio en el resultado? Se ha demostrado que la práctica distribuida o espaciada comparada con la práctica masiva o en una sola sesión, para promover la retención a largo plazo, comúnmente conocida como efecto espaciador, es más efectiva y la evidencia para ello es bastante sólida. Los intervalos más largos entre sesiones o prácticas tienen mayores efectos que los intervalos más cortos, porque al espaciar la práctica y aumentar la frecuencia de estudio esto ocasiona una mejor retención de información (Larsen et al., 2008; Yang et al., 2019).

De igual manera, se ha encontrado una relación positiva entre la mayor frecuencia de evaluaciones y el rendimiento académico. En general, se ha informado que el aumento de la frecuencia de evaluación se correlaciona positivamente con tasas más bajas de reprobación en los cursos, puntajes totales más altos en los cursos y en las evaluaciones intermedias, así como con una mayor motivación académica (Healy et al., 2017, citados en Clair y Jensen, 2020).

Ansiedad ante los exámenes

La preocupación o angustia ante los exámenes es común en el estudiantado, esta se asocia típicamente con un rendimiento deficiente, conductas de evitación, pérdida de motivación, disminución en la recuperación de la memoria y deterioro de la atención (Wolf y Smith, 1995; Zeidner, 2005; citados en Clair y Jensen, 2020). No obstante, investigaciones informan una disminución de la ansiedad ante los exámenes cuando se utiliza el aprendizaje potenciado por la evaluación, principalmente a través de cuestionarios de bajo riesgo como una forma de preparación para los exámenes finales o sumativos (Messineo et al. 2015; Nyroos et al., 2016).

Autores como Choi y Lee (2020), confirman que los beneficios de las pruebas intermedias son arrasadores y apuntan a que ayudan a disminuir distractores mentales y la ansiedad, con ello se aumenta el desempeño en el examen final, lo que facilita evitar asociaciones negativas usualmente ligadas a las pruebas de alto impacto.

En este marco de ideas en donde la evaluación es un agente poderoso para favorecer mayores niveles de logro en los aprendizajes de los estudiantes, el *aprendizaje potenciado por la prueba* es un tema prometedor que cubre los propósitos principales de la evaluación formativa como herramienta que mejora el aprendizaje de los estudiantes y que se complementa con la evaluación sumativa.

Visto de este modo, reconocer los resultados de las investigaciones puede traer grandes beneficios en el proceso educativo. La evidencia hasta aquí expuesta tiene la finalidad de enriquecer el aprendizaje de los estudiantes considerando que los conocimientos y habilidades cognitivas requeridas a través de los exámenes se encuentran alineados con los resultados de aprendizaje designados en el curso.

¿CUÁLES SON LOS BENEFICIOS DE APLICAR EL APRENDIZAJE POTENCIADO POR LA EVALUACIÓN?

Como se ha visto en las secciones anteriores, los exámenes o pruebas activan la recuperación de información, lo que a su vez propicia la retención del material de estudio en la memoria a largo plazo y con ello ganancias en el aprendizaje. Pero, ¿qué otros efectos positivos, directos o indirectos, trae consigo este fenómeno?

A continuación, se enlistan algunos de ellos con el objetivo de considerar su implementación en el espacio educativo:

- Mejora de la retención a largo plazo y la capacidad de recordar información aprendida. Esto es, cada vez que se recupera información se vuelve más fuerte y fácil de acceder en el futuro (Larsen et al., 2008).
- Produce una mejor organización del conocimiento, fortalece hábitos de estudio, disciplina y autoconfianza.
- Motiva a que los alumnos estudien más, integrar más pruebas o exámenes a lo largo del semestre alentará a repasar de manera más consistente durante el curso, por lo que se espera que el rendimiento aumente.

- Ayuda tanto a los docentes como a los estudiantes a identificar conceptos o temas en los que puedan presentar dificultades. Realizar una prueba permite reconocer lo que se sabe y lo que no, de modo que se puedan concentrar los esfuerzos de estudio en áreas en las que los conocimientos son deficientes (Roediger et al., 2011).
- Apoya a los estudiantes a identificar en qué contenidos necesitan concentrarse más cuando estudian por sí mismos.
- Favorece el rendimiento en pruebas finales o de alto impacto, al optimizar la retención, el aprendizaje en general y los resultados finales, además, la retroalimentación en estas pruebas juega un papel importante porque ayuda a los estudiantes a fortalecer su comprensión y minimizar el exceso de confianza (Appelrouth et al., 2017).
- Mejora el proceso metacognitivo de los estudiantes en tanto les permite identificar lo que saben y así, guiar las actividades de estudios posteriores.
- Cuando los estudiantes presentan un examen y posteriormente estudian de nuevo el material, aprenden más que si únicamente repasaran sin ser evaluados.
- Perfecciona la transferencia de conocimientos a nuevos contextos. Los beneficios mnemotécnicos al ser evaluados no se limitan a las preguntas o hechos específicos. La transferencia puede definirse como la aplicación del conocimiento aprendido de una situación a otra, los investigadores a menudo la clasifican como cercana si la nueva situación es similar a la de aprendizaje, y lejana si existen diferencias entre ambas (Roediger et al., 2011).
- Proporciona a los profesores información valiosa de lo que saben o no los alumnos, a su vez, pueden sugerir de manera individual o grupal distintas actividades de estudio. Si bien, en lo cotidiano, los exámenes y cuestionarios permiten evaluar el conocimiento de los estudiantes, las tareas también se pueden utilizar para este propósito.
- Desvincula las connotaciones negativas de las pruebas al visualizarlas como estrategias de estudio que mejoran el aprendizaje, incluso más que otro tipo de estrategias que los estudiantes y profesores emplean de forma más habitual.
- La evidencia sugiere efectos de las pruebas de práctica tanto en los resultados de la prueba final como para la transferencia de conocimiento, apuntando a que las pruebas de práctica se construyan para promover habilidades de pensamiento de orden superior que sean transferibles (Adesope et al., 2017).

En las actividades cotidianas, el profesorado debe tomar decisiones pedagógicas de largo alcance orientadas al logro de los propósitos de aprendizaje previamente establecidos, de allí que requieran de una brújula para orientar dichas acciones; la evaluación formativa se puede convertir en esa guía para adaptar sus estrategias e intervenciones. Sin embargo, no se debe olvidar que, por sus características, la evaluación formativa es un tema desafiante, ya que requiere de una planificación estratégica, así como del manejo de diversas metodologías y la capacidad para interpretar adecuadamente la evidencia recuperada y también se sabe de diferentes factores personales (educación, capacitación, actitud, creencias sobre la enseñanza, destreza y capacidad) o contextuales (ambiente escolar, apoyos, condiciones de trabajo, características de los estudiantes) que subyacen a la intención del docente sobre su uso (Yan et al., 2021).

Pese a la cantidad de aspectos que están involucrados, este tipo de evaluación tiene un impacto doble, por un lado, busca mejorar los aprendizajes de los estudiantes y, por otro, plantea la posibilidad de modificar o ajustar las acciones de los profesores en función de los resultados identificados (UNESCO, 2021). Con todo, prestar atención al desempeño de los estudiantes a través del aprendizaje potenciado por la evaluación, como una buena práctica de evaluación formativa, también ofrecerá a cada docente la posibilidad de obtener información para orientar su instrucción.

¿CÓMO PUEDE EL PROFESORADO IMPLEMENTARLO EN LAS CLASES?

En este apartado se presentan sugerencias sobre cómo incorporar el aprendizaje potenciado por la evaluación dentro y fuera del entorno escolar (Tabla 2).

Es recomendable que los y las docentes se aseguren de proporcionar suficientes oportunidades a los estudiantes para probar realmente el conocimiento, más que solo informarles sobre el formato o existencia de una prueba final. Considere las siguientes recomendaciones:

- *Preguntas a lo largo de la clase*. Cuestionar y esperar que los estudiantes respondan, sea de manera voluntaria o al azar, además de realimentar las respuestas correctas o no, resulta un ejercicio –aplicado por un sinnúmero de docentes– que merece especial atención cuando se habla de aprendizaje potenciado por la evaluación (Yang et al., 2019).
- *Quizzes o cuestionarios frecuentes*. Estos pueden consistir en preguntas de opción múltiple o respuesta corta, administrados en lápiz y papel o en línea (para mayor referencia del tema, revisar el Capítulo 13. Quiz, de este libro). Los estudiantes al realizar cuestionarios de práctica se darán cuenta de los temas o conceptos en los que requieren mayor empeño, al mismo tiempo, se les animará a estudiar con frecuencia porque saben que serán evaluados. Por su parte, los docentes obtendrán información inmediata de aquello que cada alumno está aprendiendo. Las pruebas acumulativas son una buena opción, es decir, pruebas en las que se incluya material previo y sea posible incrementar poco a poco el nivel de complejidad. También puede solicitar a los estudiantes argumentar la elección de la respuesta y expresar los motivos por los que no eligieron el resto de las opciones (Yang et al., 2019).
- *Recuperación en espacios educativos*. En esta actividad, solicite al estudiantado uno o más de los siguientes aspectos:

    - Mencionar al menos dos aspectos clave revisados en la última clase.
    - Señalar palabras clave que caracterizarían el contenido de la última sesión.
    - Explicar un concepto clave de la última lección.
    - Elaborar y responder, en parejas, tres o más preguntas basadas en el contenido visto hasta el momento.

- *Puntos resumen*. Hacia el final de la clase pida a los estudiantes que redacten los puntos principales de la clase, lo anterior, sin revisar sus notas de trabajo. En un estudio de

Lyle y Crawford (2011, citados en Brame y Biel, 2015) se examinaron los efectos de esta actividad y observaron un efecto significativo en la memoria de los alumnos al final del semestre.

- *Escribir o dibujar todo lo que se sabe*. Pida a sus alumnos que guarden sus materiales de estudio, posteriormente, solicite que escriban o dibujen de manera detallada lo que vieron en la clase. Al final de la actividad, verifique las producciones junto con los estudiantes para que analicen qué tan precisos fueron y cuáles son los puntos importantes que omitieron.
- *Cuadrículas de desafío o recuperación* (para mayor detalle, lo invitamos a consultar *retrieval grids* por su denominación en inglés). Resulta una estrategia útil para recapitular y revisar el contenido de la asignatura, a partir de una gama de preguntas que varían en nivel de dificultad. El contenido evaluado puede ser de la última lección, de la semana pasada e incluso de tiempo atrás. Cada pregunta cuenta con un puntaje específico, este aumenta según el nivel de dificultad.
- *Tarjetas de memoria (flashcards)*. Con la finalidad de que los estudiantes continúen repasando los temas o conceptos revisados en clase, invítelos a elaborar tarjetas de memoria o *flashcards.* Es conveniente pedirles que además de recordar la información que contienen, vaya más allá de las definiciones, estableciendo vínculos entre sus ideas. Como recurso tanto para profesores como alumnos, la guía titulada *Be Your Own Teacher: How to Study with Flashcards* puede ser de gran utilidad (https://www.learningscientists.org/blog/2016/2/20-1).
- *Práctica distribuida*. Como se vio en el apartado de hallazgos de la investigación, resulta conveniente comentar con los alumnos el efecto de "estudiar en el último momento" o espaciar las sesiones de aprendizaje. Por ejemplo, invítelos a estudiar una hora por día o repasar un tema durante un día, luego una semana, un mes o hasta un semestre después, a partir de haberlo revisado por primera vez. Recuérdeles que "la práctica hace al maestro". Para conocer más acerca de la práctica de recuperación como experiencia de evaluación en el proceso de enseñanza y aprendizaje lo invitamos a ver el video titulado *Study Strategies: Retrieval Practice* (The Learning Scientists, 2016. https://youtu.be/Pjrqc6UMDKM).
- *Informar a los estudiantes sobre el aprendizaje potenciado por la evaluación*. Como docente puede ayudar en las habilidades metacognitivas de sus estudiantes, decirles que los exámenes pruebas o cuestionarios frecuentes ayudan a aprender, y que pueden tomar una variedad de formas, puede brindarles una herramienta particularmente útil para su caja de herramientas de aprendizaje (Stanger-Hall et al., 2011; citados en Brame y Biel, 2015). Señalar los beneficios potenciales, directos e indirectos del aprendizaje potenciado por las pruebas, puede llevar a los alumnos a mejorar su propio proceso de aprendizaje. Es importante tener cuidado de no generar la idea de que el estudio está enfocado a obtener un mejor resultado en el examen final o sumativo, "enseñar para el examen", sino por el contrario, que se están promoviendo estrategias de estudio para fortalecer el aprendizaje y la transferencia de este conocimiento a nuevas y diversas situaciones fuera de los espacios educativos.

Esta lista no pretende ser exhaustiva, es un punto de partida desde el cuerpo de conocimiento de la psicología cognitiva para implementar los principios del aprendizaje potenciado por la evaluación, principalmente a través de exámenes o ejercicios de bajo riesgo. Sin embargo, la aplicación dependerá de cada clase y contexto escolar.

CONCLUSIONES

Desde una visión formativa e iterativa, las acciones pedagógicas y la graduación o modificación de estas buscan alcanzar los propósitos de aprendizaje, para ello el aprendizaje potenciado por la evaluación es una manera de optimizar el conocimiento de los estudiantes. Dada la diversidad de contextos educativos, este capítulo pretende dar un panorama general de los beneficios directos e indirectos de las pruebas frente a otras formas de estudio. De hecho, aunque los estudiantes que solo vuelven a estudiar el material predicen que recordarán mejor la información en una semana que aquellos a quienes se les evalúa, los resultados empíricos demuestran exactamente lo contrario (Adesope et al., 2017; Collins et al., 2019; Roediger y Karpicke, 2006).

El aprendizaje potenciado por la evaluación, principalmente a través de exámenes o pruebas, activa procesos de recuperación de información, lo que a su vez favorece el aprendizaje del material estudiado, el almacenamiento en la memoria a largo plazo y el monitoreo de las estrategias de estudio de los alumnos. En este sentido, resulta oportuno considerar la utilización de este enfoque como una práctica para promover el aprendizaje y no solo como una herramienta de evaluación sumativa, pues desde la evidencia académica publicada, ayuda a las funciones cognitivas básicas como la recuperación de memoria, y también a procesos de mayor complejidad como el que se requiere para el aprendizaje inductivo que prepara a los estudiantes para la adquisición de material totalmente nuevo o la aplicación de conocimiento, ya que entre más profundo es el procesamiento requerido en la prueba de práctica, el desempeño en futuras pruebas es mejor (Choi y Lee, 2020; Collins et al., 2018).

Cabe recordar que en la actualidad se deben establecer mecanismos de aprendizaje en doble vía, por un lado, fortalecer las habilidades básicas, pero a la vez, promover procesos cognitivos complejos, de pensamiento crítico, argumentación y creación que requiere un mundo cada vez más desafiante (UNESCO, 2021). Así, los beneficios del aprendizaje potenciado por la evaluación se dan independientemente de los niveles educativos y contextos escolares. Con ello, los profesores pueden considerar la práctica de recuperación dentro de la planificación de actividades. Lo anterior, no equivale a crear una saturación de exámenes, también se puede incorporar el estudio autodirigido, dentro y fuera del espacio educativo.

Si a lo largo del curso, los profesores hacen preguntas a los estudiantes durante las clases; les piden que expliquen “con sus propias palabras” lo que se revisó en la sesión anterior; aplican exámenes o cuestionarios; emplean las pruebas como una forma de aplicar lo que se aprendió, entonces, probablemente están fomentando dentro del proceso de enseñanza

y aprendizaje la recuperación, retención y consolidación de la información a largo plazo. Además, si proporcionan retroalimentación para guiar el desempeño, pueden mejorar los resultados obtenidos.

Compartir con los estudiantes el potencial de las pruebas o exámenes para el aprendizaje, los conducirá a una mejor elección de estrategias de estudio que promuevan el rendimiento a largo plazo. Hay que reiterar que *"a medida que usamos nuestros recuerdos, las cosas que recordamos se vuelven más memorables. Mientras que las cosas que compiten con los recuerdos se vuelven menos memorables"* (Stachowiak, 2015).

Finalmente, con esta información, se invita al docente a repensar y replantear el uso de las pruebas y otras formas de práctica de recuperación que podrían usarse para promover el aprendizaje, ya que cualquier estrategia o método que demuestre, a través de evidencia empírica, mejoras para los estudiantes, debe ser analizada para identificar su valor en los espacios educativos reales y, en su caso, adoptada por las comunidades educativas para su puesta en práctica en las actividades cotidianas. Lo anterior, a fin de realizar cambios en los paradigmas y prácticas habituales para la optimización de los resultados de aprendizaje en los estudiantes.

REFERENCIAS

Adesope, O., Trevisan D. y Sundararaja, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659-701.

Appelrouth, J., Zabrucky, K. y Moore, D. (2017). Preparing students for college admissions tests. *Assessment in Education: Principles, Policy & Practice, 24*(1), 78-95. https://doi.org/10.1080/0969594X.2015.1075958

Butler, A. C. y Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36*(3), 604–616. https://link.springer.com/content/pdf/10.3758/MC.36.3.604.pdf

Brame, C. y Biel, R. (2015). Test-enhanced learning: The potential for testing to promote greater learning in undergraduate science courses. *Life Sciences Education, 14*(2), 1-12. https://www.lifescied.org/doi/full/10.1187/cbe.14-11-0208

Choi, H. y Lee, H. (2020). Knowing is not half the batle: the role of actual test experience in the forward testing effect. *Educational Psychology Review,* 32, 765-789. https://doi.org/10.1007/s10648-020-09518-0

Collins, D., Rasco, D. y Benassi, V. (2018). Test-enhanced learning: Does deeper processing on quizzes benefit exam performance? *Teaching of Psychology, 45*(3), 235-238. https://doi.org/10.1177/0098628318779262

Clair, B. y Jensen, J. (2020). Modulators of test-enhanced learning in post-secondary Biology. *Bioscene: Journal of College Biology Teaching, 46*(2), 22-29. https://files.eric.ed.gov/fulltext/EJ1285488.pdf

Green, M., Moeller, J. y Spak, J. (2018). Test-enhanced learning in health professions education: A systematic review: BEME Guide No. 48. *Medical Teacher, (40)*4, 337-350. https://doi.org/10.1080/0142159X.2018.1430354

Greving, S. y Richter, T. (2008). Examining the testing effect in university teaching: Retrievability and question format matter. *Frontiers in Psychology, 9,* 1-10. https://doi.org/10.3389/fpsyg.2018.02412

Larsen, D., Butler, A. y Roediger, H. (2008). Test-enhanced learning in medical education. *Medical Education, 42*(10), 959-966. https://doi.org/10.1111/j.1365-2923.2008.03124.x

Little, J. y Bjork, E. (2015). Optimizing multiple choice tests as tools for learning. *Memory & Cognition, 43*(1), 14–26. https://bjorklab.psych.ucla.edu/wp-content/uploads/sites/13/2017/01/LittleBjorkMC2014.pdf

McDaniel, M., Roediger, H. y McDermott, K. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14*(2), 200-206. https://link.springer.com/content/pdf/10.3758/BF03194052.pdf

Messineo, L., Gentile, M. y Allegra, M. (2015). Test-enhanced learning: analysis of an experience with undergraduate nursing students. *BMC Medical Education, 15*(182), 1-7. https://bmcmededuc.biomedcentral.com/articles/10.1186/s12909-015-0464-5

Nyroos, M., Schéle, I. y Wiklund-Hörnqvist, C. (2016). Implementing test enhanced learning: Swedish teacher students' perception of quizzing. *International Journal of Higher Education, 5*(4), 1-12. https://www.sciedu.ca/journal/index.php/ijhe/article/view/10210

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO]. (2021). *Evaluación formativa: Una oportunidad para transformar la educación en tiempos de pandemia. Reflexión a partir de los resultados del estudio cualitativo sobre perspectivas docentes en torno a la evaluación formativa.* OREALC/UNESCO Santiago. https://unesdoc.unesco.org/ark:/48223/pf0000378045

Roediger, H. y Karpicke, J. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*(3), 181-210. https://doi-org.pbidi.unam.mx:2443/10.1111/j.1745-6916.2006.00012.x

Roediger, H., Putnam, A. y Smith, M. (2011). Ten benefits of testing and their applications to educational practice. En J. P. Mestre & B. H. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (pp. 1–36). Elsevier Academic Press. https://www.researchgate.net/publication/291166244_Ten_Benefits_of_Testing_and_Their_Applications_to_Educational_Practice

Stachowiak, B. (Anfitriona). (2015, 29 de octubre). How to use cognitive psychology to enhance leare ning with Robert Bjork [episodio de podcast de audio]. En *Teaching in Higher Ed podcast*. https://teachinginhighered.com/podcast/cognitive-psychology/

The Learning Scientists. (2016, 19 de septiembre). *Study strategies: Retrieval practice* [Video]. YouTube. https://www.youtube.com/watch?v=Pjrqc6UMDKM

Widiastuti, I., Mukminatien, N., Prayogo, J. y Irawati, E. (2020). Dissonances between teachers' beliefs and practices of formative assessment in EFL classes. *International Journal of Instruction, 13*(1), 71-84. https://doi.org/10.29333/iji.2020.1315a

Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L. y Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*. https://doi.org/10.1080/0969594X.2021.1884042

Yang, B., Razo, J. y Persky, A. (2019). Using testing as a learning tool. *American Journal of Pharmaceutical Education, 83*(9). 1862-1872. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6920642/

Capítulo 8

EL RETO DEL ESTABLECIMIENTO DE ESTÁNDARES DE EVALUACIÓN Y PUNTOS DE CORTE EN EDUCACIÓN SUPERIOR

Mildred López, Gabriela González

INTRODUCCIÓN

La alineación entre el currículo educativo diseñado, los contenidos abordados en el proceso enseñanza-aprendizaje, y la evaluación determina la percepción de que el proceso ha sido efectivo y justo. La evaluación es sin duda uno de los momentos más críticos en la vida estudiantil de aprendices y docentes, es la *hora de la verdad*, el momento donde se demuestra la efectividad del aprendizaje. Si el proceso de aprendizaje ha sido un círculo virtuoso, este es, en sí mismo, un momento de aprendizaje en el que el estudiante recibe evidencias de su desarrollo y sugerencias de áreas o estrategias que aún necesita trabajar. Cuando este proceso deja mucho que desear, el momento está cargado de emociones como miedo, frustración y enojo.

Si bien, la primera capa de análisis es importante, es decir el análisis en lo individual por parte de los estudiantes, para reflexionar sobre qué puedo hacer mejor; el análisis que debemos hacer como educadores y diseñadores de programas también tiene que estar presente. Este análisis exige que la evaluación sea por diseño de alta calidad, que garantice el egreso de estudiantes con el más alto estándar, y que los instrumentos y procesos que utiliza para clasificar a los estudiantes, hayan sido conceptualizados con el mismo nivel de exigencia.

Los objetivos de este capítulo son:

- Analizar los conceptos de establecimiento de estándares en evaluación y puntos de corte.
- Contrastar los principales criterios que se utilizan para el establecimiento de puntos de corte en la definición de estándares.
- Discutir las implicaciones para el aprendizaje a distancia resultado de la pandemia del nuevo coronavirus.

DESARROLLO DEL TEMA

Una definición quizás un tanto simplista del concepto de evaluación es que esta fase del proceso enseñanza-aprendizaje colecta evidencia del desempeño del estudiante. Es posible describir un momento de evaluación como un instante en el tiempo en el que se toma una fotografía al desempeño de una persona. Esta fotografía puede ser tan favorecedora u horrible tanto como el escenario, la iluminación y la escenografía hayan sido dispuestas. También tiene mucho que ver la experiencia del modelo para saber cómo posar y del fotógrafo para gestionar esos elementos en el plató. Siguiendo con la alegoría de las fotografías, para conocer verdaderamente qué es capaz de hacer el *modelo* en la imagen, es importante tener acceso a la mayor cantidad de fotografías posibles donde demuestre su desempeño en una amplitud de contextos, estilos y actitudes. En la evaluación del aprendizaje, es igual. Solo que las fotografías son en realidad los resultados de distintos instrumentos y pruebas que se obtienen en diferentes etapas del desarrollo profesional. En la medida que se tienen más *fotografías*, más se tiene certeza de que la valoración realizada fue correcta.

Boud (2006) afirma que la evaluación genera más ansiedad en los estudiantes e irritación en los docentes que cualquier otro elemento de la educación superior. Para los estudiantes significa quizás horas de estudio adicional el fin de semana, tratar de memorizar cientos de páginas de libros de texto y diapositivas de una presentación. Su preocupación gira al menos, en torno a dos elementos: el primero se refiere al momento donde estarán en el examen o la prueba en la que tienen un tiempo limitado para responder preguntas, y el segundo al juicio que se hará sobre su desempeño con este instrumento. Para los docentes, esta ansiedad se encuentra en el diseño de las pruebas, las fechas límite para entregar reactivos y el estrés de enfrentarse a estudiantes decepcionados o iracundos una vez que se entreguen los resultados (Joughin, 2008).

¿Qué es el establecimiento de estándares?

Cuando la evaluación se utiliza además para la toma de decisiones, diversas inferencias surgen a partir de las calificaciones o la respuesta de los evaluados. Estas inferencias tienen grandes repercusiones ya que determinan por ejemplo qué desempeño es *suficientemente bueno* para ser admitido a una universidad o cuánto es suficiente saber de matemáticas para avanzar al siguiente año escolar. Una definición sencilla sobre el establecimiento de estándares es que se refiere al proceso de decidir qué es lo *suficientemente* bueno (Cusimano, 1996). Barman (2008) añade a la discusión diciendo que el establecimiento de estándares es una actividad de generación de políticas, en un nivel más conceptual fija un ideal, mientras que la operacionalización de este ideal es la puntuación de corte descrita en dicha política. Para fines prácticos, el reto en la definición de un estándar se refiere en realidad a la definición de un punto de corte, es decir el nivel mínimo aprobatorio que separa entre desempeño competente del que no lo es.

En modelos educativos basados en competencias que describen un perfil de egreso, la promesa es que cada graduado domina todas estas competencias en un nivel aceptable. Cada competencia es evaluada usando el mismo criterio para todas y todos los graduados de la misma generación y es muy similar a las generaciones anteriores. Esto quiere decir que los

resultados de las evaluaciones deben separar en dos grupos, excluyentes entre sí, aquellos que se conforman al estándar definido, que tienen un desempeño competente y los que no, quienes aún deben trabajar para alcanzar un nivel de desempeño aceptable. Esta evaluación basada en criterios valora el desempeño del participante en sí mismo, independientemente del desempeño de otros miembros del grupo. Es decir que todos los estudiantes que cumplan con los criterios mínimos obtendrán un documento que avala su competencia, ya sea un título o diploma de grado.

Aunque esta definición de absolutos parece intuitiva de entender, cualquier persona que haya puesto un pie en un salón de clases y haya interactuado con estudiantes sabe que una evaluación absoluta, única y objetiva en un solo momento de interacción es imposible de obtener. A decir verdad, el desempeño de los estudiantes varía de ser razonablemente competente, competente acompañado de un docente, o muy competente, por nombrar algunos. Este problema se agrava al traducir un desempeño a una calificación, ¿cómo se compara la competencia de un estudiante que obtuvo un 80% con el que obtuvo un 90% en una escala donde el 70% es el corte aceptable para definir el pase en esta prueba? Muchas de estas pequeñas variaciones son subjetivas y difícilmente traducible a criterios observables.

Figura 1: Pasos para establecimiento de estándares

Elección del método

Integración de equipo de jueces o expertos en el área de conocimiento valorada

Definición del participante limítrofe

Capacitación de jueces o expertos en el método elegido

Establecimiento de puntos de corte por parte de jueces o expertos

Otro caso de definición de estándares que frecuentemente enfrentamos en educación superior, son aquellos estándares en los que el desempeño de un participante será comparado con otros que sustentaron evaluación al mismo momento, o con algunos que lo han tomado con anterioridad. Por ejemplo, en este caso, no todos los participantes que sustentan un examen o se esforzaron *mucho* en un programa educativo, tendrán acceso a una beca, o una plaza de

posgrado. He ahí el compromiso de que estas definiciones de estándares y las mismas valoraciones de la prueba tengan una definición, conceptual o empírica, avalada por expertos y un análisis de su validez.

Una adecuada definición de estándares debería ser sensible a distinguir el desempeño de los participantes, tener sustento estadístico, ser confiable y fácil de implementar y aplicar por cualquier persona (Barman, 2008). En los últimos años, diversos autores han generado alternativas para desarrollar estos estándares y sustentar las decisiones que estas conllevan, estos varían en cómo deben llevarse a cabo estos juicios de valor, quién debe participar en emitir los juicios, y en qué información deben basarse.

Sin importar el método utilizado para fijar estos estándares y establecimiento de puntos de corte, el proceso debe contener al menos seis pasos (Ruano et al., 2018), descritos en la Figura 1.

La selección del método depende de los siguientes elementos (Ruano et al., 2018):

- Experiencia previa con el método: utilizar un método con el que se está familiarizado reduce la cantidad de tiempo utilizado en los pasos de la Figura 1, además que el equipo de diseño se siente confiado del proceso que está llevando a cabo.
- Tiempo disponible para la definición de estándares: impacta en la elección de métodos centrados en la prueba que pueden ser valorados por un grupo de jueces o expertos considerando un grupo hipotético, o si es posible realizar un piloto para tener los resultados reales de participantes para los cuales fue diseñada la prueba.
- Tipo de preguntas utilizadas: algunos métodos son recomendados para pruebas con reactivos de opción múltiple, otros de respuestas dicotómicas o preguntas abiertas.
- Validez del método: diversos autores han levantado inquietudes respecto a la validez de las decisiones realizadas por distintos métodos. Por ejemplo, existen cuestionamientos sobre el método de grupos contrastados y el Angoff, las cuales son abordadas más adelante.
- Además de considerar la definición de estándares desde el punto de vista metodológico, es importante que el desarrollo de la evaluación sea analizado como algo integral en el que los resultados sean útiles, y defendibles ante los estudiantes, profesores, desarrolladores de la prueba y demás grupos de interés (Cetin y Gelbal, 2013).

Principales métodos para fijar estándares

Históricamente, han existido numerosas propuestas para fijar dichos estándares, tan solo en 1986 existían cerca de 38 métodos para definir estándares, diez años más tarde en 1996 existían al menos 50 (Cusimano, 1996). De acuerdo con el propósito de la evaluación, los métodos para fijar estándares se pueden clasificar en dos tipos, aquellos centrados en la prueba y los centrados en la persona. El primero se enfoca en que los expertos provean un estimado basándose en los reactivos de una evaluación. El segundo se enfoca en la observación del desempeño de los grupos, por lo que debe tener revisiones a lo largo del tiempo (Downing et al., 2010). La Tabla 1 presenta un vistazo a los principales métodos de acuerdo con esta categorización.

Tabla 1: Métodos centrados en la prueba vs en la persona

Enfoque	Estándar
Centrados en la prueba	Angoff.
	Nedelsky.
	Bookmark o de marcador.
	Ebel.
Centrados en la persona	Grupos contrastados.
	Borderline o métodos de frontera.

En los métodos basados en los criterios de la prueba, el contenido del examen es revisado por un grupo de jueces expertos, entre los que destacan el método de Angoff, el Nedelsky, el método de marcador y Ebel. Mientras que en los métodos basados en la persona se concentran en evaluar el desempeño del sujeto, algunos métodos son el de grupos contrastados y los métodos de frontera. Aunque existen diferentes métodos y variaciones de estos, cada método tiene sus ventajas y desventajas lo que los hace más adecuados para una u otra aplicación (Cetin y Gelbal, 2013). A continuación, se describen algunas de las más utilizadas.

MÉTODO DE ANGOFF

Este es uno de los métodos de evaluación estándar más populares y tiene una larga historia de éxito debido a que las calificaciones son fácilmente obtenidas, los cálculos son simples y el método es fácil de comunicar. Su propuesta nace en los años 70's, Angoff presentó un método sistemático para decidir la calificación mínima para aprobar una evaluación. El método sugiere que previo a aplicar una prueba, se lleve a cabo una revisión de todas las preguntas con un grupo de jueces expertos en el área, llamados panelistas, como lo pueden ser los mismos profesores, o invitados externos y sin ninguna relación con la institución (Papageorgiou y Tannenbaum, 2016).

En esta evaluación, los jueces establecen un puntaje entre el 0 y 100% dependiendo de la dificultad de la pregunta. Se le da un puntaje cercano a 100% a aquellas preguntas en la evaluación que se consideran que una persona calificada podría contestar correctamente y cercano a 0 a las cuales se consideran que el evaluado contestaría de manera incorrecta. La suma de todos los puntos obtenidos es considerada el punto de corte (Zieky, 2001). En la Tabla 2 se presenta un ejemplo donde se tienen 4 panelistas que gradúan la dificultad de 10 preguntas. Posteriormente se estima un punto de corte por cada panelista, y finalmente se calcula el promedio de los puntos de corte estimados para fijar cuál sería este.

Tabla 2: Ejemplo de método Angoff

Pregunta	Panelista 1	Panelista 2	Panelista 3	Panelista 4
1	0.90	0.85	0.75	0.80
2	0.65	0.65	0.60	0.55
3	0.85	0.90	0.85	0.80
4	0.65	0.60	0.65	0.60
5	0.55	0.50	0.45	0.45
6	0.60	0.80	0.85	0.90
7	0.75	0.70	0.70	0.85
8	0.80	0.90	0.75	0.90
9	0.65	0.55	0.50	0.50
10	0.80	0.95	0.90	0.95
Punto de corte por panelista	7.20	7.40	7.00	7.30
Promedio de punto de corte			7.23	

Para 1990, el método Angoff continuaba como el método de elección para establecer un punto de corte, era ampliamente utilizado y estudiado; sin embargo, fue atacado debido a que expertos consideraban que este y todos los métodos basados en la evaluación de las preguntas eran poco factibles de desarrollar. Algunos de los argumentos más fuertes en contra de estos métodos son: la alta inversión de costo y tiempo, además de esto, el ensamblar un panel de expertos no es nada fácil, ya que se debe asegurar una representatividad que impactará la definición del estándar. Una vez armado, algunos de los retos son asegurar que los panelistas sean capaces de hacer lo que este método requiere, y estimar la probabilidad de que un estudiante hipotético conteste correctamente (Katz y Tannenbaum, 2014). No obstante, es común que el estándar haya sido fijado muy alto, por lo que se han propuesto distintas modificaciones en la que los panelistas tienen permitido ajustar sus estimados después de la *dosis de realidad* de ver las respuestas de los participantes (Schoonheim-Klein et al. 2009).

Por otro lado, diversos estudios han reportado implementaciones donde los jueces involucrados en las revisiones de diferentes evaluaciones se sentían confiados con su habilidad para realizar las tareas del método Angoff, entendían que tenían que hacer y cómo hacerlo y creían que los resultados eran defendibles y suficientemente buenos (Papageorgiou y Tannenbaum, 2016). Este proceso para establecer un estándar es utilizado en la actualidad en diferentes pruebas de lenguaje como lo es el examen de inglés como segundo idioma (TOEFL), que es una prueba estandarizada de dominio del idioma inglés. Algunas innovaciones que han surgido se enmarcan en la inclusión de tecnología para llevar a cabo las sesiones con panelistas en la virtualidad y hacer el análisis de forma más eficiente (Katz y Tannenbaum, 2014).

MÉTODO NEDELSKY

En 1954, Leo Nedelsky sugirió un método de evaluación en el cual se determinan estándares de calificación absolutos para pruebas objetivas, específicamente aquellas de opción múltiple. Este método se volvió muy popular no solo en exámenes escolares sino también en certificaciones profesionales y licenciaturas (Chang, 2009). Nedelsky se basaba en la idea de conceptualizar a los estudiantes en el límite entre aprobar y reprobar (llamándolos los estudiantes F-D), e identificar las opciones de las preguntas de opción múltiple que estos estudiantes serían capaces de eliminar como incorrectas. El número de opciones remanentes es la probabilidad de que el estudiante mínimamente competente conteste correctamente. La suma de las probabilidades se considera el puntaje esperado para este estudiante y es la base para el punto de corte.

El método de Nedelsky es popular en contextos médicos, presuntamente porque el evaluado, debe de ser capaz de rechazar esas opciones que “causarían daño a un paciente”. A lo largo del método de Nedelsky existen dos suposiciones que podrían influir en los resultados, las cuales son, que los jueces sean capaces de discriminar entre las opciones que los evaluados pudieran considerar como opción y aquellas que no, y que los evaluados pueden responder correctamente contestando de manera aleatoria (Zieky, 2001).

Un ejemplo del método de Nedelsky se muestra en la Tabla 3, la cual muestra una evaluación de una prueba con 10 preguntas de opción múltiple, en ella el juez descarta las opciones prediciendo que una persona mínimamente competente sería capaz de identificar esas respuestas como incorrectas. Por lo tanto, en la segunda pregunta la probabilidad sería de 0.50, la recomendación de este juez sobre el punto de corte es de 4.83.

Tabla 3: Ejemplo de método Nedelsky

Pregunta	Opciones					Restantes	Probabilidad
1	~~A~~	~~B~~	C	~~D~~	~~E~~	1	1.00
2	A	B	~~C~~	~~D~~	~~E~~	2	0.50
3	~~A~~	~~B~~	C	D	E	3	0.33
4	A	B	C	~~D~~	E	4	0.25
5	A	B	C	D	~~E~~	4	0.25
6	A	~~B~~	~~C~~	D	E	3	0.33
7	~~A~~	~~B~~	~~C~~	~~D~~	E	1	1.00
8	A	~~B~~	~~C~~	~~D~~	E	2	0.50
9	~~A~~	~~B~~	C	D	E	3	0.33
10	A	B	~~C~~	~~D~~	E	3	0.33
Recomendación que este juez hace sobre el punto de corte							4.83

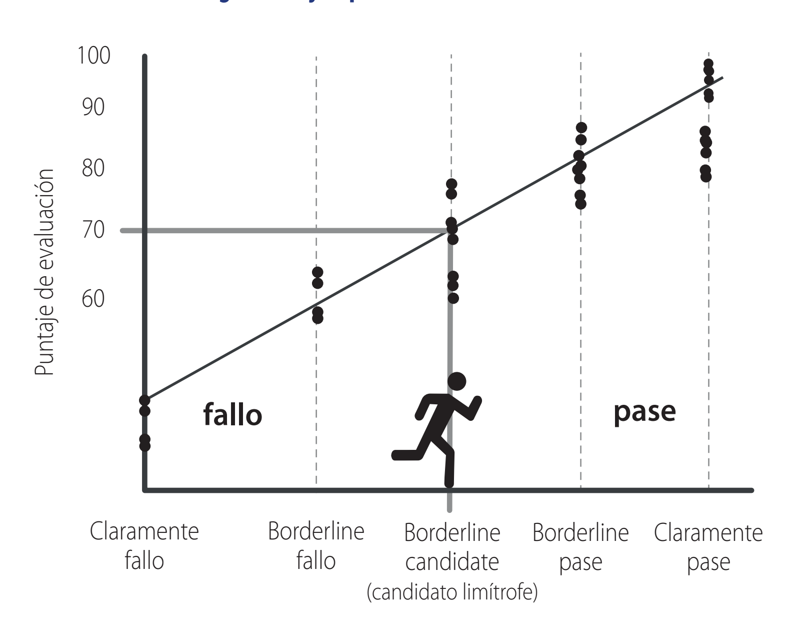
BORDERLINE GROUP

El método Borderline Group es un método centrado en el examinado en lugar de un método centrado en la prueba, por lo que los juicios se hacen sobre los examinados individuales, no sobre el contenido del examen. El método se puede utilizar solo cuando los expertos en contenido que están calificados para servir como emisores de estándares, como lo son los profesores o expertos, observan de manera directa la prueba de desempeño. Las calificaciones globales de los jueces observadores se utilizan para determinar la puntuación de la lista de verificación que se utilizará como estándar de aprobación.

Este método, como el de Nedelsky y Angoff, se basa en la premisa de que la calificación aprobatoria debe ser aquella obtenida por cualquier evaluado mínimamente competente; sin embargo, en lugar de evaluar cada una de las preguntas u opciones, los jueces evalúan directamente a quienes tomen la prueba y posteriormente, la calificación media se toma como calificación aprobatoria (Schoonheim-Klein, 2009).

El procedimiento para llevar a cabo este método inicia mediante la preparación de los jueces orientándolos en el caso y los instrumentos para calificar. Posteriormente los jueces observan directamente el rendimiento de todos los evaluados. Como se muestra en la Figura 2, los jueces proveen una calificación global del rendimiento general de cada uno en una escala de tres puntos, no aprobado (fallo), borderline (limítrofe) y aprobado (pase).

Figura 2: Ejemplo del Método Borderline



Una debilidad del método es que el número de examinados clasificados como dentro del límite es a menudo pequeño. En la mayoría de las situaciones, si es posible recopilar calificaciones, también será posible recopilar datos para implementar el método de grupos contras-

tantes. Si es así, es preferible contrastar grupos porque los datos están directamente relacionados con la decisión a tomar.

Implicaciones, problemas y críticas hacia los modelos existentes

Sin importar el método o modelo elegido para la definición del estándar, este involucra juicios subjetivos de expertos, análisis psicométrico, definición de políticas para la toma de decisiones y la socialización de estos resultados (Papageorgiou y Tannenbaum, 2016). Es por eso por lo que es una tarea inseparable, el definir criterios claros, justos y defendibles para que deje de ser una tarea tan controvertida (Downing et al., 2010).

Una crítica para la definición de estándares es que estos fueron fijados en ciertas condiciones históricas, sociales y culturales, asumiendo un ideal que difícilmente es alcanzable en las pruebas de examinación donde los participantes tienen mucho en juego, por ejemplo, en aquellas de ubicación en un programa de idiomas o aquellas que garantizan la entrada a una plaza de estudios de posgrado.

Otro problema altamente relacionado se refiere a la interpretación de resultados de los diferentes grupos de interés. Frecuentemente, los resultados se usan en dos niveles de interpretación, el primero alrededor de la estimación de las habilidades de los participantes en la evaluación, y el segundo para fines de evaluación de la calidad de los programas formativos. Si el foco es este, los estándares no deben ser la excusa para señalar culpables o fincar responsabilidades a estudiantes, escuelas o países, ya que los datos deben tener un análisis agregado y la reflexión enfocarse en los contextos donde estas evaluaciones están teniendo lugar.

IMPLICACIONES DE LA PANDEMIA Y EVALUACIÓN A DISTANCIA

La pandemia del COVID-19 ha traído consigo muchos retos para la educación, desde la adaptación de un programa educacional a un ambiente virtual de enseñanza, el mantenernos cerca a los estudiantes, así como el diseño de actividades y evaluaciones.

Los grandes cambios que hemos experimentado al traer la escuela o la oficina a la casa, se han traducido en la creación de ambientes compartidos donde los alumnos enfrentan la dificultad para ajustarse a un nuevo método de estudio, responsabilidades en el hogar, y distracciones como ruidos u otro tipo de interrupciones que afectan su concentración.

Aunque hablar de educación híbrida o a distancia se ha convertido en algo común en los diferentes niveles educativos, como el hecho que formará parte de una nueva normalidad incluso una vez que el virus no represente una amenaza para la salud; la calidad y la validez de una educación remota continúa siendo un tema controversial.

Particularmente, la discusión gira alrededor del constante reto de la educación, la evaluación. En primer lugar, cómo distribuirla y llevarla a cabo de manera exitosa, además de cómo asegurar la integridad o disminuir las oportunidades para deshonestidad escolar.

McCabe et al. (2012), señalan que la incidencia de deshonestidad académica ha ido en aumento en los últimos años, adjudicándose a la facilidad de acceso a fuentes electrónicas

que tienen las nuevas generaciones. Debido a esto, distintas autoridades escolares han tomado diferentes acciones para combatir la deshonestidad en pruebas académicas incluyendo el uso de la tecnología. En particular, se ha hecho popular el uso de programas y plataformas que evitan que el alumno pueda acceder a otras ventanas o bloquean el uso de internet después de haber descargado el examen. Esto resuelve los problemas para exámenes de opción múltiple enfocados en la memorización de información, pero fallan en considerar otros factores para valorar el desempeño integral del alumno.

Junto a esto se han utilizado técnicas como mantener el micrófono y cámara encendidos en todo momento, el uso de varios dispositivos para que el evaluador pueda visualizar todo el ambiente en el cual se realiza el examen e incluso la reducción del tiempo en el examen para evitar que se busque información por otros medios. Sin embargo, aunque estas técnicas pueden parecer atractivas, dejan de lado las dificultades del acceso a la tecnología que muchos de nuestros estudiantes han enfrentado. Junto a la labor de evitar que los alumnos incurran en una deshonestidad, también es importante considerar el acceso a recursos de aprendizaje, así como internet y equipos de cómputo, y los básicos como agua o electricidad, que pueden impactar en la valoración del desempeño que se hace del alumno. Definir estándares de desempeño debe ser sensible también a los contextos alrededor del diseño, implementación y análisis de las pruebas.

CONCLUSIONES Y REFLEXIONES FINALES

La definición de estándares de evaluación está ligada a los currículos basados en competencias, donde las universidades y los programas educativos deben garantizar un nivel mínimo de aptitud. Los métodos discutidos previamente fueron establecidos hace más de cincuenta años, con el fin de encontrar una forma de evaluar de manera estándar a un grupo ampliado de estudiantes. En la actualidad, el foco de la educación ha cambiado, ya no es la educación masificada, ahora el enfoque es una educación flexible y personalizada, los retos que enfrentamos han cambiado y el modo de evaluar debe evolucionar también, tomando en cuenta diferentes factores para valorar un buen desempeño.

A lo largo de la pandemia del COVID-19, los retos cambiaron y las evaluaciones de manera remota han demostrado que pueden tener altas repercusiones aun en el rendimiento de los alumnos. Estas posibles repercusiones incluyen la presión añadida de alcanzar una nota alta, por lo que algunos alumnos han sentido también la necesidad de cometer deshonestidades académicas. A pesar de que la tecnología ha facilitado el aprendizaje y ha sido la base de la educación remota, la evaluación sigue siendo un reto debido a la imposibilidad de aplicar estándares a todos los alumnos sin considerar el ambiente en casa o las disparidades en el acceso a la educación a distancia.

REFERENCIAS

Barman, A. (2008). Standard setting in student assessment: is a defensible method yet to come? *Annals Academy of Medicine Singapore, 37*(11), 958-963.

Boud, D. (2006). Foreword. En C. Bryan & K. Clegg (Eds.), Innovative assessment in higher education (pp. XVII–XIX). London and New York: Routledge.

Cetin, S., Gelbal, S. (2013). A Comparison of Bookmark and Angoff Standard Setting Methods. *Educational Sciences: Theory and Practice, 13*(4), 2169-2175.

Chang, L. (2009). Judgmental Item Analysis of the Nedelsky and Angoff Standard-Setting Methods. *Applied Measurement in Education, 12*(2), 151-165.

Cusimano, M.D. (1996). Standard Setting in Medical Education. *Academic Medicine, 71*(10), 112-121.

Downing, S. M., Tekian, A., Yudkowksy, R. (2010). Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. *Teaching and Learning in Medicine, 18*(1), 50-57.

Joughin, G. (2008). Assessment, Learning and Judgement in Higher Education: A Critical Review. En: G. Joughin (eds) Assessment, Learning and Judgement in Higher Education, 1–15.

Kampa, N., Wagner, H., Köller, O. (2019). The standard setting process: validating interpretations of stakeholders. *Large-scale Assessments in Education, 7*, 3. https://doi.org/10.1186/s40536-019-0071-8

Katz, I., Tannenbaum, R.J. (2014). Comparison of Web-based and Face-to-face Standard Setting using the Angoff method. *Journal of Testing Technology, 15*(1),1-17.

Lázaro, J.L., Usart, M., Gisbert, M. (2019). La evaluación de la competencia digital docente: construcción de un instrumento para medir los conocimientos de futuros docentes. Journal of New Approaches in Educational Research, *8*(1), 75-81.

McCabe, D. L., Butterfield, K. D., & Trevino, L. K. (2012). *Cheating in college: Why students do it and what educators can do about it*. JHU Press.

Papageorgiou, S., & Tannenbaum, R. J. (2016). Situating Standard Setting within Argument-Based Validity. Language Assessment Quarterly, *13*(2), 109–123. https://doi.org/10.1080/15434303.2016.1149857

Ruano, A.L., Vizuete, A., Moreno, J.C., Quispe, W. (2018). Habilitación profesional: caso Ecuador. Rev. Ecu. Med. Eugenio Espejo. *7*(9), 15-20.

Schoonheim-Klein, M., Muijtjens, A., Habets, L., Manogue, M., Van Der Vleuten, C., & Van Der Velden, U. (2009). Who will pass the dental OSCE? Comparison of the Angoff and the borderline regression standard setting methods. European Journal of Dental Education, *13*(3), 162–171. https://doi.org/10.1111/j.1600-0579.2008.00568.x

Zieky, M. J. (2001). So much has changed: How the setting of cutscores has evolved since the 1980s. En G. J. Cizek (Ed.), Setting performance standards: Concepts, methods, and perspectives (pp. 19–51). Lawrence Erlbaum Associates Publishers.

Capítulo 9

EVALUACIÓN EN LÍNEA

Melchor Sánchez Mendiola, Elibidú Ortega Sánchez

"No preguntes qué puede hacer la tecnología por ti,
más bien pregunta qué necesita la pedagogía."
GILLY SALMON, UNIVERSIDAD DE LEICESTER

"…los exámenes en línea solo se deben considerar cuando se hayan examinado exhaustivamente los temas de acceso igualitario a la infraestructura y la conectividad, la seguridad y los métodos de supervisión en línea, la transparencia, y las habilidades y brechas digitales de los estudiantes."
UNESCO, 2020

INTRODUCCIÓN

Uno de los desafíos más importantes en la educación formal es la evaluación, especialmente en épocas marcadas por crisis. En el año 2020, la pandemia global por el virus SARS-CoV-2 que causa la enfermedad COVID-19 afectó seriamente la vida de estudiantes y profesores al interrumpir las clases presenciales y mantener cerradas las escuelas por periodos prolongados (Burgos, D., Tlili A. y Tabacco, A., 2021; UNESCO, 2021). Súbitamente los profesores se vieron forzados, sin una planeación intencionada, a transferir los contenidos, materiales didácticos y métodos e instrumentos de evaluación de sus asignaturas de un formato presencial a plataformas de aprendizaje en línea. Ello implicó una lucha en varios frentes para proporcionar plataformas digitales con la capacidad suficiente para atender a estudiantes y docentes, al tiempo que todos enfrentaban variaciones en la calidad de acceso a Internet, en condiciones heterogéneas y ante una brecha digital creciente. Los múltiples aspectos tradicionalmente complicados de la evaluación del aprendizaje, convergieron durante el inicio de la pandemia en lo que García-Peñalvo denominó "evaluación online: la tormenta perfecta" (García-Peñalvo, 2020a).

Aún en momentos de crisis no todo debe cambiar, por ejemplo, la evaluación del y para el aprendizaje presencial o a distancia, requiere obtener evidencias para conocer los logros de aprendizaje de los alumnos o sus necesidades de apoyo, así como la definición de estrategias de evaluación y selección de los instrumentos adecuados para dicho propósito. Basados en las experiencias de evaluación y aprendizaje mediados por tecnología durante la pandemia, este capítulo tiene como objetivos proporcionar información útil al profesorado sobre lo que debe tomarse en cuenta en la evaluación en línea y compartir algunas prácticas y lecciones aprendidas en la Universidad Nacional Autónoma de México (UNAM) en este campo.

El capítulo aborda: a) qué es la evaluación en línea; b) particularidades, ventajas y desventajas de esta modalidad, b) recomendaciones para la planeación, diseño, aplicación y análisis de resultados de la evaluación en línea, c) una experiencia en la UNAM y d) conclusiones y reflexiones finales.

¿QUÉ ES LA EVALUACIÓN EN LÍNEA?

La educación en línea o mediada por tecnología tiene una larga historia en educación superior. En las últimas décadas ha tenido un progreso espectacular, convirtiéndose en una modalidad cada vez más utilizada en universidades de todo el mundo (Singh y Thurman, 2019). A pesar de los enormes avances en las herramientas tecnológicas de hardware y software, el uso educativo centrado en el aprendizaje de los estudiantes no ha avanzado al mismo ritmo que los dispositivos y plataformas digitales disponibles. Lamentablemente los principios básicos de la buena evaluación del y para el aprendizaje no han permeado totalmente en el uso cotidiano de estas sofisticadas herramientas. La pandemia exhibió crudamente la falta de una integración apropiada de la tecnología con las actividades docentes del profesorado y de aprendizaje de los estudiantes, mientras que empujó el uso masivo de las herramientas digitales para realizar exámenes a distancia, incluso en casa, con todas las amenazas a la validez que ello implica (Burgos, D., Tlili A. y Tabacco, A., 2021; UNESCO, 2020, 2021).

De la misma forma que la comunidad académica internacional continúa discutiendo la definición precisa de los términos utilizados en educación en línea (Singh y Thurman, 2019), no hay un acuerdo global sobre las definiciones precisas de la evaluación en línea. Para fines de este capítulo, podemos entender a la educación en línea ("*online learning*" o "*e-learning*") como un término paraguas que se refiere al aprendizaje apoyado por las tecnologías de información y comunicación (TIC), que en la actualidad generalmente requiere acceso a Internet por medio de dispositivos digitales (Moore et al., 2011; Singh y Thurman, 2019). De la gran cantidad de términos que se han usado para describir las facetas de este concepto (aprendizaje en línea, educación mediada por tecnología, educación a distancia, aprendizaje híbrido, entre muchos otros), se han desprendido varias etiquetas a sus aspectos evaluativos: evaluación en línea, evaluación asistida por computadora, *e-assessment*, exámenes por computadora, evaluación digital, evaluación electrónica, entre otros (Cantillon et al., 2004; Dennick et al, 2009; Northcote, 2008).

Para fines del capítulo nos referimos a la evaluación en línea como el uso de las tecnologías de información y comunicación (que incluye dispositivos digitales, software y plataformas virtuales, redes de conectividad alámbrica e inalámbrica) para realizar actividades de evaluación del, para y como aprendizaje, con énfasis en la promoción del aprendizaje del estudiantado (capítulos 1 y 5 de este libro). La evaluación mediada por tecnología es de particular importancia en la actualidad, y debe llevarse a cabo siguiendo los principios fundamentales de la evaluación descritos en el capítulo 1 de este libro, en virtud de que "…los estudiantes pueden, con dificultad, escapar de los efectos de una pobre enseñanza, pero no pueden escapar de los efectos de una pobre evaluación" (Boud, 1995).

VENTAJAS Y DESVENTAJAS DE LA EVALUACIÓN EN LÍNEA

La evaluación en línea en educación tiene una serie de características que han evolucionado a diferentes ritmos, dependiendo del contexto socioeconómico y de la velocidad de adopción de innovaciones en las universidades y organizaciones dedicadas a la evaluación. Como todo lo que implica el uso de la tecnología, hay factores humanos, culturales, sociales, psicológicos, tecnológicos, económicos y del contexto que determinan el uso de estas herramientas. En general, el avance de la evaluación en línea antes de la pandemia, como una herramienta cotidiana para promover el aprendizaje de los estudiantes, había sido lento, con excepción de los exámenes sumativos de alto impacto en universidades u organizaciones dedicadas a ello (*Educational Testing Service*, *National Board of Medical Examiners*, entre otros). En estos casos el uso de herramientas digitales para generar exámenes, la metodología de aplicación en centros *ad hoc*, el desarrollo de bancos de reactivos, análisis psicométricos, mecanismos de seguridad e identificación de los sustentantes, procesamiento natural de lenguaje y uso de inteligencia artificial, tuvieron importantes avances (Dennick et al., 2009; Doğan et al., 2020; Hols-Elders et al., 2008; JISC, 2010;).

De cualquier manera, diversos autores han identificado grupos de ventajas y desventajas desde el punto de vista administrativo y educativo, para los diferentes actores del proceso de evaluación (instituciones, docentes y estudiantes) (Cantillon et al, 2004; Conrad y Openo, 2018; Dennick et al., 2009). Por supuesto que algo que constituye una ventaja en un contexto para una persona o institución, puede ser también una desventaja en otro contexto o momento, por lo que el siguiente listado no es exclusivo ni limitativo a cada actor. A continuación, se describen algunos de los pros y contras más relevantes, por población implicada:

DOCENTES

Pros

- Ahorra tiempo para calificar exámenes y analizar resultados.
- Pueden seguir el "rastro digital" del desempeño de los estudiantes individual y grupalmente, visualizar y archivar datos con facilidad.
- Puede disminuir el uso de tiempo de profesores para vigilar personalmente aplicaciones de exámenes.
- La información de evaluación de los estudiantes y retroalimentación de sus cursos se recibe con mayor eficiencia.
- Los datos de evaluación recabados longitudinalmente en los cursos pueden integrarse en la misma plataforma virtual, con el resto de los componentes de los cursos y asignaturas.
- Facilita proporcionar retroalimentación y evaluación formativa individual y grupal las 24 horas, los 7 días de la semana.
- Permite usar más variedad de medios (gráficas, imágenes, videos, audios) y de tipos de reactivos, comparado con exámenes en papel.

- Mayor posibilidad de control de calidad con el uso de las herramientas que proveen las plataformas.
- Automatización de varios elementos del proceso de evaluación (calificación, promedios, gráficas, etc.)
- Disponibilidad de varias herramientas para detectar plagio en las tareas y exámenes, así como minimizar trampa y conductas deshonestas en las evaluaciones.
- Permite realizar exámenes adaptativos en línea (siempre y cuando se tenga un robusto banco de reactivos), con la consecuente eficiencia de tiempo de los estudiantes al responder exámenes.
- Facilidad de tener acceso a la plataforma de evaluación desde cualquier dispositivo con acceso a la red, a cualquier hora del día.
- Disponibilidad de una gran cantidad de apps y plataformas digitales para evaluación, varias de bajo costo o gratuitas, que permiten un proceso evaluativo más dinámico e interactivo, así como evaluar el trabajo colaborativo.
- Facilita explorar otros métodos e instrumentos de evaluación, como auto-evaluación, evaluación por pares, exámenes de libro abierto, evaluación auténtica, simuladores y realidad virtual.

Contras

- Requiere habilidades digitales, que implican tiempo, esfuerzo y recursos, así como participación en actividades de formación docente.
- Implica tener dispositivos digitales y licencias de software, que tienen un costo para el profesorado o para la institución.
- La disponibilidad permanente puede saturar el tiempo del profesorado e invadir su agenda personal y de descanso, generando desgaste y fatiga.
- Favorece exámenes objetivos de opción múltiple, limitando las opciones de tipos de evaluación (aunque están surgiendo herramientas de procesamiento natural del lenguaje y análisis del discurso que pueden revolucionar la evaluación en línea).
- Requiere disciplina y trabajo constante para mantener actualizados los bancos de reactivos y los archivos.
- Puede introducir amenazas a la validez, al dar ventajas a estudiantes de mayor nivel socioeconómico o que tengan mayores habilidades digitales y acceso a Internet, así como dispositivos más poderosos.
- Puede hacer más difícil vigilar estudiantes en centros de aplicación de exámenes sumativos, si las pantallas ocultan al estudiante y las computadoras están muy cerca una de la otra.
- Existe la percepción de que es más fácil hacer trampa y copiar, usar herramientas de ayuda no autorizadas o comunicarse virtualmente con otras personas.
- Se pueden convertir en “cajas negras” para los profesores, al desconocer estos el funcionamiento interno de los algoritmos del software.

- La enorme cantidad de opciones de apps y plataformas para evaluación hace difícil la elección de las más apropiadas para el contexto local, y el recambio continuo y actualizaciones de estas herramientas complican el mantenerse al día en su uso eficaz.
- Puede generar la percepción de alejamiento personal con los estudiantes, provocando la deshumanización de la práctica docente.

ESTUDIANTES

Pros

- Pueden recibir retroalimentación instantánea en su desempeño y mantener comunicación constante con profesores y compañeros.
- En general los estudiantes están familiarizados con el uso de la tecnología, y se sienten confiados al usarla.
- Les permite monitorear su progreso académico durante el curso.
- Tienen acceso ilimitado y permanente a los recursos en línea del curso y de la universidad.
- Prefieren el uso de multimedios (audio, video, redes sociales), mayor interactividad y experiencia visual.
- Las herramientas tienen múltiples opciones de accesibilidad para los estudiantes con necesidades especiales.
- Al responder exámenes pueden modificar respuestas fácilmente, y tener el "mapa" del examen y preguntas que faltan, así como etiquetar aquellas en las que tienen dudas.
- Se disminuye la separación tradicional entre evaluación y enseñanza, al enfrentarse a plataformas digitales que integran contenidos, experiencias y tareas con herramientas evaluativas como cuestionarios y quizzes.

Contras

- Implica que los estudiantes tengan dispositivos digitales y licencias de software, que tienen un costo para ellos o para la institución.
- Requiere acceso a Internet en los espacios educativos o en casa, con sus consecuentes retos y costos.
- Varias herramientas digitales no están diseñadas para los teléfonos inteligentes, que son el principal dispositivo de los estudiantes.
- Al desconocer los retos del análisis psicométrico de los exámenes y de los aspectos técnicos de la evaluación, pueden exigir resultados inmediatos en los exámenes sumativos de alto impacto, creando amenazas a la validez.
- La gran cantidad y heterogeneidad de aplicaciones, plataformas y sistemas operativos usados por sus diferentes maestros, crean confusión y complican la vida de los estudiantes.
- En general los estudiantes están menos familiarizados con los aspectos asincrónicos de la evaluación en línea formativa y longitudinal, pueden no responder adecuadamente a la retroalimentación en ambientes virtuales.

- Riesgo de invasiones a la privacidad y uso inadecuado de datos personales, por compañeros, docentes o personal institucional.
- Posibilidad de privar a los alumnos del contacto personal cara-a-cara con los docentes.

INSTITUCIONES

Pros

- Ahorro de tiempo para analizar estadísticamente los resultados y automatizar diversos procesos de los exámenes.
- Pueden aplicarse exámenes simultáneamente en diferentes sitios.
- Ahorro de dinero y recursos en el uso de papel, más amable con el ambiente.
- Centralización de los apoyos a la docencia y exámenes, generando economías de escala.
- Generación y mantenimiento de bancos de reactivos y plataformas de exámenes en línea para toda la institución, fácilmente actualizables.
- Permite atender a números grandes de usuarios y realizar evaluación a gran escala.
- Posibilidad de tener una política adecuada de gobernanza de datos, para su uso adecuado y ético, siguiendo las normativas nacionales e internacionales.
- Disminución del potencial sesgo y de errores en la calificación manual de exámenes (aunque el creciente reto de los sesgos incluidos en los algoritmos de inteligencia artificial no se ha resuelto).

Contras

- Requiere una inversión inicial importante en adquisición de equipo y software, así como de constante mantenimiento, actualizaciones y apoyo técnico.
- Requiere sistemas muy confiables y seguros, por el riesgo de fallas eléctricas, de los dispositivos, o de hackers externos y virus. La ciberseguridad es un problema creciente.
- Prevalencia alta de una cultura de piratería de software.
- Mayor facilidad de que ocurra plagio y problemas de derechos de autor de material (imágenes, audio, documentos).
- Necesidad de mantener un equipo de personal técnico y administrativo para dar sostén y apoyo técnico a estos procesos.
- Inversión en servidores físicos y virtuales para el enorme caudal de recursos digitales y productos que se generan en la educación y evaluación en línea. Necesidad de mantenimiento constante y espacios físicos acondicionados.
- Necesidad de inversión en infraestructura de conectividad física e inalámbrica en todas las instalaciones.
- Difícil planear y usar puntualmente los presupuestos anuales, debido a la rápida evolución y elevados precios de la tecnología. La obsolescencia de los productos en pocos años es un gran reto para renovarlos.

- Necesidad de proveer capacitación continua y apoyo técnico a todos los usuarios, administradores, docentes y estudiantes.
- Complejidad de autentificación de los sustentantes de los exámenes, para eliminar el riesgo de suplantación se requiere tecnología sofisticada.

Como es aparente en este listado de ventajas y desventajas de la evaluación en línea y mediada por la tecnología, la cantidad de factores a considerar en su diseño, planeación e implementación son múltiples y requieren de las acciones concertadas y consensuadas de diversos actores del proceso educativo. Si bien la pandemia motivó una adopción urgente y masiva de estas metodologías, su uso continuado y progresivo debe ser fruto de una amplia y profunda deliberación que tome en cuenta todos los aspectos arriba señalados, para asegurar que la evaluación cumpla su función de promover el aprendizaje profundo de los estudiantes, disminuyendo la carga de trabajo para el profesorado, a un costo razonable para todos los involucrados.

RECOMENDACIONES PARA LA EVALUACIÓN EN LÍNEA

Las recomendaciones para evaluación mediada por tecnología se pueden presentar de acuerdo con los tres momentos principales que conlleva cualquier proceso de evaluación, y que corresponden a su planeación y diseño (antes), aplicación (durante) y análisis de resultados (después). Para la elaboración de estas recomendaciones se tomaron en cuenta diversas publicaciones recientes sobre el tema (JISC, 2010; Padayachee et al., 2019; Rahim, 2020; Sánchez-Mendiola, 2021; The International Test Commission; 2006; UNESCO, 2021; Zhou, 2016).

1) **Planeación y diseño.** Este tipo de evaluación demanda una etapa de planeación en la que se establezca claramente qué se desea evaluar, para qué y cómo (¿cuáles son los instrumentos disponibles para ello y cuál es su evidencia de validez?). La selección de instrumentos cobra especial relevancia en la aplicación en línea porque es necesario considerar si el contenido de la asignatura, módulo o curso es teórico o práctico y debe buscarse una plataforma que sirva de medio adecuado para realizar una evaluación sincrónica o asincrónica.

   Consejo 1. Hay literalmente cientos de instrumentos para evaluar cada una de las habilidades cognitivas y no cognitivas de los seres humanos, por lo que dependiendo de lo que se requiera hay que elegir y usar el o los instrumentos pertinentes.

   Las recomendaciones previas a la aplicación de un instrumento de evaluación son (Rahim, 2020):

- **Valorar prerrequisitos para implementar la evaluación en línea**. Sobre los factores institucionales son indispensables el tipo recursos, políticas y prácticas; sobre los factores docentes hay que valorar la necesidad de capacitarse en prácticas de evaluación en

línea y en el uso de las TIC; y sobre los factores de estudiantes hay que tener conocimiento sobre cuáles son los recursos tecnológicos y de acceso a internet con los que cuentan.

Consejo 2. Recuerde proteger los contenidos digitales sensibles, respetar las normas de privacidad y derechos de autor. Para ello, es importante comprender el uso y la creación de licencias abiertas y recursos educativos abiertos, incluyendo su correcta atribución.

- **Asegurar la alineación de la evaluación con los objetivos de aprendizaje.** Identifique los resultados de aprendizaje más importantes de la asignatura, módulo o curso para la evaluación. Revise con atención cómo fueron enseñados a distancia y trate de hacer la evaluación congruente con el contenido y con la forma de enseñanza. Asegúrese de que el contenido que refleje los aprendizajes esperados sea susceptible de ser evaluado a distancia.
- **Tomar en cuenta la diversidad de situaciones de los estudiantes.** Estas situaciones van desde las tecnológicas como el acceso a Internet, hasta otras como el desempleo, combinar el trabajo de casa con el estudio.
- **Mantener un buen equilibrio entre evaluación formativa y sumativa.** Diversificar los instrumentos de evaluación, nos permite tener un conjunto de mediciones que facilitan realizar una retroalimentación efectiva y, por otra parte, permiten al final otorgar una calificación de carácter sumativo.

Consejo 3. Promueva evaluaciones formativas que requieran interacción y colaboración entre los estudiantes.

- **Establecer una clara comunicación con los estudiantes respecto a los criterios de evaluación.** Es deseable que los alumnos conozcan muy bien qué se espera de ellos, cuáles van a ser los instrumentos de evaluación, para que ellos puedan tener más control y sentirse más seguros en el momento de la evaluación. Esto evita dudas o quejas y hace más transparente el proceso para todos los involucrados. No deje de informarles el día, la hora, los procedimientos y las guías para la evaluación.
- **Diseñar la logística de aplicación.** Se deben considerar aspectos tales como identificación de los sustentantes, los sistemas de supervisión electrónica o *proctoring*, tiempo de resolución y soporte técnico durante la aplicación del instrumento de evaluación.

2) **Aplicación.** Después de identificar las estrategias e instrumentos a utilizar, es necesario "aplicarlos con prudencia, sensatez y adecuarlos al contexto en el proceso de aprendizaje de los estudiantes" (Zhou K., 2016).

- **Considerar el formato, horario y tiempo de resolución de los exámenes**. La aplicación del examen debe ser oportuna y factible, para ello es importante considerar los tiempos de otras asignaturas y las demás actividades de los estudiantes.

- **Apegarse a la logística de aplicación diseñada.** Verificar la identidad de los sustentantes, monitorear la aplicación de las evaluaciones en tiempo real, dar soporte técnico en caso de ser necesario, cerrar la plataforma en el tiempo indicado para la aplicación y considerar el tiempo efectivo de aplicación para aquellos estudiantes que tuvieron problemas técnicos durante la aplicación.
- **Tener en cuenta que la tecnología solo es un instrumento.** Es fundamental tener conciencia de que todas las plataformas digitales y dispositivos solo son herramientas, el instrumento más importante en la evaluación del y para el aprendizaje sigue siendo el docente. La interpretación juiciosa de los resultados de las evaluaciones formativas y sumativas, así como su uso prudente y humano en el proceso educativo, sigue siendo prerrogativa de los seres humanos.

3) **Análisis de los resultados.** El análisis de resultados es similar al que se lleva a cabo de manera presencial, solo que en un contexto en línea es necesario considerar si la aplicación fue síncrona o asíncrona, si la retroalimentación va a ser grupal o individual y a través de qué medio. Las recomendaciones para este rubro son las siguientes:

- **Estimular el aprendizaje de los estudiantes con evaluación en línea**. En tiempos de crisis es muy importante mantener el ánimo de los estudiantes, relacionarnos más con ellos para generar un ambiente de confianza. De esta manera cuando reciben retroalimentación y sus calificaciones, pueden encontrar oportunidades de mejora y ver el lado positivo del proceso. Si por el contrario la evaluación es cuesta arriba, no le encuentran sentido y lo viven como un castigo, va a ser difícil que concluyan el curso con éxito.
- **Garantizar una retroalimentación de alta calidad**. La retroalimentación es fundamental para mantener el interés de los alumnos en el contenido de las asignaturas. Los hace sentir acompañados, acogidos y comprendidos. A través de una buena retroalimentación podemos identificar puntos positivos y aspectos a mejorar en las clases, así como hacer algo al respecto.
- **Identificar las amenazas a la validez de la evaluación.** Reconocer que los resultados de las evaluaciones pueden ser afectados por una subrepresentación del constructo o por factores que ocasionen varianza irrelevante al constructo, nos permite identificar el grado en que se puede confiar en la interpretación correcta de los resultados obtenidos.

Consejo 4. Una evaluación sin validez en las inferencias que se hagan sobre sus resultados, no es útil y puede ser peligrosa porque se pueden tomar decisiones sin fundamento.

Al final del día, la evaluación en línea debe satisfacer los mismos principios fundamentales de cualquier evaluación del, para y como aprendizaje, así como proporcionar los elementos necesarios para que docentes, estudiantes e instituciones educativas logren un balance sistémico en el proceso, que privilegie el aprendizaje de los estudiantes. Un modelo que puede propiciar la adecuada realización de la evaluación en línea, es el marco conceptual de

los cuatro pilares de apoyo para esta modalidad propuestos por Padayachee y cols. (2019), esquematizados en la Figura 1.

Figura 1. Marco de los cuatro pilares de apoyo para la evaluación en línea (AEL) (adaptado de Padayachee et al, 2019).

Evaluación en línea

ENTORNO DE APRENDIZAJE FÍSICO	ENTORNO DE APRENDIZAJE INTELECTUAL	ENTORNO DE APRENDIZAJE EMOCIONAL	ENTORNO DE APRENDIZAJE VIRTUAL
• Arreglo espacial. • Trae su propio dispositivo. • Evaluaciones compatibles con dispositivos móviles. • Accesibilidad para capacidades diferentes. • Variedad de recursos. • Personas participantes. • Soporte técnico. • Capacitación. • Habilidades de mediación para aprendizaje híbrido. • Tiempo.	• Retroalimentación. • Prácticas tutoriales. • Aprendizaje independiente y autoevaluación. • Cultura de aprendizaje. • Recordatorio de actividades. • Monitorear aprendizaje propio. • Participación activa en el aprendizaje. • Aprendizaje colaborativo. • Metas y objetivos explícitos. • Clarificar los resultados desde el inicio. • Tiempo.	• Niveles de ansiedad bajos. • Confianza y seguridad propia. • Adopcion de la tecnología. • Confidencialidad. • Responsabilidad del aprendizaje. • Actitud positiva hacia el aprendizaje. • Establecimiento de expectativas altas. • Capacitación y apoyo. • Proporcionar un código de conducta para la comunicación en línea. • Netiqueta. • Democratización de la experiencia de aprendizaje. • Tiempo.	• Montaje de tecnología. • Especializaciones de la evaluación en línea. • Diseño. • Oportunidades para entrenamiento, práctica, comunicación en línea. • Dimensión social. • Literacidad digital. • Facilidad de uso. • Recursos en línea. • Tutores en línea. • Comunidad de aprendizaje. • Mantener estudiantes comprometidos y conectados. • Tiempo.

Ambiente de apoyo al aprendizaje

Aprendizaje mixto | Dominio del aprendizaje

EXPERIENCIAS DE EVALUACIONES EN LÍNEA EN LA DIRECCIÓN DE EVALUACIÓN EDUCATIVA (DEE) DE LA UNAM

La intención de este apartado es mostrar a través de un ejemplo, cómo se transforma la logística de aplicación de un examen presencial comparada con la situación a distancia en línea. El contenido se retoma de cómo se aplicó el Examen Diagnóstico de Ingreso al Bachillerato de la UNAM, elaborado en la Dirección de Evaluación Educativa (DEE) antes y durante la pandemia (Tabla 1).

Tabla 1. Comparación de logística de aplicación del Examen Diagnóstico de Ingreso al Bachillerato de la UNAM, en la aplicación del examen presencial frente a la aplicación en línea en casa.

Aplicación Presencial (papel y lápiz)	Aplicación A distancia (en línea)
Antes del examen	
1. Elaborar la versión del examen. 2. Contactar a las entidades académicas del bachillerato de la UNAM para acordar fechas y horarios de aplicación. 3. Capacitar a las entidades académicas en donde se aplicará el examen de manera presencial. 4. Cada plantel del bachillerato: ● Determina el medio por cual informará a los alumnos fechas y horarios de acceso al examen, así como los espacios físicos para realizar la aplicación. ● Calcula la cantidad de estudiantes que presentarán el examen para solicitar los ejemplares a imprenta por medio de la DEE. ● Reciben en el plantel las cajas con los cuadernillos y hojas de respuesta necesarios para la aplicación.	1. Elaborar la versión del examen. 2. Contactar a las entidades académicas del bachillerato de la UNAM para acordar fechas y horarios de aplicación. 3. Capacitar a las entidades académicas que aplicarán el examen en línea. 4. Cada plantel del bachillerato: ● Determina el medio por cual informará a los alumnos fechas y horarios de acceso al examen, así como las características del equipo de cómputo con el que deberán de contar para realizarlo. ● Nombra un enlace técnico para recibir las posibles dudas de los alumnos al momento de realizar el examen y en caso necesario hacerlas llegar a la DEE. ● Dan a conocer las condiciones de ética y confidencialidad que deben cumplir los sustentantes durante el examen, así como las recomendaciones generales para los alumnos.
Durante el examen	
5. La aplicación del examen es monitoreada por personal del plantel en tiempo real con dos profesores por grupo. 6. Si los examinados tienen alguna duda es resuelta por los monitores. 7. Los examinados tienen acceso todo el tiempo a su cuadernillo y hoja de respuesta. 8. El tiempo de resolución del examen es controlado por los monitores y es de tres horas. 9. Al concluir los responsables de aplicación reempaquetan los cuadernillos y hoja de respuestas para su regreso a la DEE, separando el material utilizado del no utilizado.	5. La aplicación del examen es monitoreada por el sistema en tiempo real. 6. El personal de sistemas de la DEE da soporte a las sedes en caso de ser necesario, a través de un correo electrónico habilitado para uso exclusivo de la aplicación. 7. El acceso al sistema se cierra una vez que concluye el tiempo indicado para la aplicación del examen. 8. El tiempo efectivo del examen comienza a partir de que se contesta la primera pregunta. La duración de la aplicación es de tres horas. 9. Al concluir el examinado hace clic en "finalizar examen" y sus resultados son guardados en la plataforma del examen.
Después del examen	
10. Realizar la clasificación de las hojas por plantel y proceder a su lectura. 11. Preparar las bases de datos del examen para realizar el análisis estadístico correspondiente para su calificación y la emisión de resultados (limpieza de bases). 12. Entregar los resultados del examen a las autoridades correspondientes en formato físico y en electrónico.	13. Exportar la base de datos de la plataforma. 14. Preparar las bases de datos del examen para realizar el análisis estadístico correspondiente para su calificación y la emisión de resultados. 15. Entregar los resultados del examen a las autoridades correspondientes en formato electrónico.

Durante la pandemia las instituciones educativas y organizaciones que realizan evaluaciones del aprendizaje han publicado gran cantidad de experiencias que pueden ser de utilidad para los docentes de todos los niveles (Burgos et al., 2021; Chen et al., 2021; Guangul et al., 2020; Holden et al., 2021; UNESCO, 2020, 2021). En la UNAM se elaboraron una serie de recomendaciones para la docencia y la evaluación en línea, que está disponibles gratuitamente en la red, con un listado de consejos basados en la evidencia científica publicada, ligas de sitios de interés, bibliografía anotada y listado de herramientas digitales para la evaluación (Sánchez-Mendiola et al., 2021). En España varios académicos y universidades publicaron guías sobre la misma temática (Abella García et al., 2020; Cabero-Almenara et al., 2021; García-Peñalvo et al., 2020b, 2021).

Por otra parte, se incrementó sustancialmente la literatura académica sobre la evaluación en línea, con reportes de casos, trabajos de investigación, revisiones narrativas y revisiones sistemáticas sobre este complejo tema (Alexandr y Mirela, 2020; Butler-Henderson y Crawford, 2020; Cavalcanti et al., 2021; Doğan, et al., 2020).

CONCLUSIONES Y REFLEXIONES FINALES

Las evaluaciones del aprendizaje, tanto las formativas como las sumativas, deben ser analizadas y, en su caso, reconsideradas y replanteadas en la era moderna, como consecuencia de la evolución de la educación en línea e híbrida, así como los efectos de las crisis globales. Se sugiere que los criterios y procedimientos de evaluación que se ajusten para fines sumativos, se lleven a cabo en un marco de alta sensibilidad hacia los estudiantes, que consideren las distintas circunstancias de los alumnos (acceso a medios y recursos digitales, situación contextual personal y familiar de salud, económica y afectiva). Sobre las evaluaciones realizadas se recomienda que sean las esenciales, es decir, aquellas que tengan la validez suficiente para evidenciar los logros de aprendizaje del alumnado. Es fundamental no saturar a los estudiantes con demandas excesivas de actividades y tareas, que los conduzcan a realizar trabajos superficiales, de escaso valor y que no contribuyan a su formación. (Sánchez-Mendiola et al., 2021)

En las aplicaciones de evaluaciones mediadas por tecnología, uno de los problemas importantes es la necesidad de conocer y garantizar la identidad de la persona examinada, así como controlar el contexto físico en el que la persona está realizando el examen. Una posible solución a estos problemas proviene de los sistemas de supervisión electrónica o *proctoring* (Carstairs y Myors, 2009; González-González et al. en Burgos, D., Tlili A. y Tabacco, A., 2021; Okada et al., 2019). Sin embargo, sería un error considerar que toda evaluación en línea debe estar condicionada a estos sistemas de supervisión electrónica, ya que tienen profundas implicaciones éticas y técnicas que deben tomarse en cuenta.

A continuación, se describen algunas conclusiones:

- Los principios básicos y fundamentales de la evaluación del, para y como aprendizaje deben aplicarse de forma similar en la evaluación en línea, tomando en cuenta las particularidades de esta modalidad y el contexto en el que se realiza la evaluación. La validez de las inferencias que se realicen en la evaluación en línea es tan importante como en la presencial.
- La evaluación en línea mediada por tecnología tiene diversas ventajas y desventajas, que deben tomarse en cuenta en las diferentes etapas del proceso de evaluación (planeación, implementación, análisis y seguimiento). A pesar de lo impresionante de los avances tecnológicos y de sus grandes ventajas y beneficios, existen riesgos sustanciales y costos que deben ser considerados para minimizarlos.
- La evaluación mediada por tecnología ofrece posibilidades de proporcionar retroalimentación rápida y efectiva para los estudiantes, debe usarse para mejorar el proceso de aprendizaje.
- Los peligros y complicaciones de la evaluación en línea ocurren principalmente en las evaluaciones sumativas de alto impacto y gran escala, como se describe en el capítulo 5 de esta obra.
- Los avances en tecnología, ciencias de la computación, pedagogía, analítica del aprendizaje, procesamiento natural del lenguaje e inteligencia artificial, tienen gran potencial para revolucionar el campo de la evaluación en línea, integrándola con los métodos de enseñanza modernos.

REFERENCIAS

Abella García, V., Grande de Prado, M., García-Peñalvo, F. J., y Corell, A. (2020). *Guía de recomendaciones para la evaluación online en las Universidades Públicas de Castilla y León. Versión 1.1*. Castilla y León, España: Universidad de Burgos, Universidad de León, Universidad de Salamanca y Universidad de Valladolid. Tomado de: https://bit.ly/2SqTtR2.

Alexandr, A., & Mirela, M. (2020) When old becomes new: a case study of oral examination as an online assessment tool, *Assessment & Evaluation in Higher Education*, *45*(8), 1205-1221, DOI: 10.1080/02602938.2020.1730301

Boud D. 1995. Enhancing learning through self assessment. London: Kogan Page.

Burgos, D., Tlili A. y Tabacco, A. (2021). *Radical Solutions for Education in a Crisis Context. COVID-19 as an Opportunity for Global Learning*. Singapore: Springer. https://link.springer.com/content/pdf/10.1007%2F978-981-15-7869-4.pdf

Butler-Henderson, K., & Crawford, J. (2020). A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Computers & education,* 159, 104024. https://doi.org/10.1016/j.compedu.2020.104024

Cabero-Almenara, J., y Palacios-Rodríguez, A. (2021). La evaluación de la educación virtual: las e-actividades. *RIED. Revista Iberoamericana De Educación a Distancia*, *24*(2), 169–188. https://doi.org/10.5944/ried.24.2.28994

Cantillon, P., Irish, B., & Sales, D. (2004). Using computers for assessment in medicine. *BMJ (Clinical research ed.)*, *329*(7466), 606–609. https://doi.org/10.1136/bmj.329.7466.606

Carstairs, J., & Myors, B. (2009). Internet testing: A natural experiment reveals test score inflation on a high-stakes, unproctored cognitive test. *Computers in Human Behavior, 25*(3), 738–742. https://doi.org/10.1016/j.chb.2009.01.011

Cavalcanti Anderson, P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y., Gašević, D., Ferreira Mello, R. (2021). Automatic feedback in online learning environments: A systematic literature review, *Computers and Education: Artificial Intelligence*, *2*, 2021, 100027, https://doi.org/10.1016/j.caeai.2021.100027. https://www.sciencedirect.com/science/article/pii/S2666920X21000217)

Chen, Z., Jiao, J., & Hu, K. (2021). Formative Assessment as an Online Instruction Intervention: Student Engagement, Outcomes, and Perceptions. *International Journal of Distance Education Technologies (IJDET), 19*(1), 50-65. http://doi.org/10.4018/IJDET.20210101.oa1

Conrad, D. L., & Openo J. (2018). Assessment strategies for online learning: engagement and authenticity. Edmonton Canada: AU Press, Athabasca University. https://www.aupress.ca/app/uploads/120279_99Z_Conrad_Openo_2018-Assessment_Strategies_for_Online_Learning.pdf

Dennick, R., Wilkinson, S., & Purcell, N. (2009). Online eAssessment: AMEE guide no. 39. *Medical Teacher*, *31*(3), 192–206. https://doi.org/10.1080/01421590902792406

Doğan, N., Kıbrıslıoğlu Uysal, N., Kelecioğlu, H., & Hambleton, R. K. (2020). An overview of e-assessment. *Hacettepe University Journal of Education, 35* (Special Issue), 1-5. doi: 10.16986/HUJE.2020063669 http://www.efdergi.hacettepe.edu.tr/yonetim/plugins/uploads/files/3700-published.pdf

García-Peñalvo, F. J. (2020a). Evaluación online: la tormenta perfecta. Recuperado de: https://bit.ly/2yO3K39.

García-Peñalvo, F. J., Corell, A., Abella-García, V., y Grande, M. (2020b). La evaluación online en la educación superior en tiempos de la COVID-19. *Education in the Knowledge Society, 21,* 12. https://doi.org/10.14201/eks.23013

García-Peñalvo F.J., Corell A., Abella-García V., Grande-de-Prado M. (2021) Recommendations for Mandatory Online Assessment in Higher Education During the COVID-19 Pandemic. In: Burgos D., Tlili A., Tabacco A. (eds) *Radical Solutions for Education in a Crisis Context. Lecture Notes in Educational Technology.* Springer, Singapore. https://doi.org/10.1007/978-981-15-7869-4_6

Guangul, F.M., Suhail, A.H., Khalit, M.I. *et al.* Challenges of remote assessment in higher education in the context of COVID-19: a case study of Middle East College. *Educ Asse Eva Acc 32***,** 519–535 (2020). https://doi.org/10.1007/s11092-020-09340-w

Holden, O. L., Norris, Meghan, E., Kuhlmeier, V. A. (2021). Academic Integrity in Online Assessment: A Research Review. *Frontiers in Education*, 6:258 https://www.frontiersin.org/article/10.3389/feduc.2021.639814

Hols-Elders, W., Bloemendaal, P., Bos, N., Quaak, M., Sijstermans, R., & De Jong, P. (2008). Twelve tips for computer-based assessment in medical education. *Medical Teacher*, *30*(7), 673–678. https://doi.org/10.1080/01421590802279595

JISC [Joint Information Systems Committee]. (2010). *Effective Assessment in a Digital Age. A guide to technology-enhanced assessment and feedback.* Higher Education Funding Council for

England (HEFCE) on behalf of JISC. https://facultyinnovate.utexas.edu/sites/default/files/digiassass_eada.pdf

Moore, J.L., Dickson-Deane, C., Galyen, K. (2011). e-Learning, online learning, and distance learning environments: Are they the same? *The Internet and Higher Education*, *14*(2), 129-135, https://doi.org/10.1016/j.iheduc.2010.10.001. (https://www.sciencedirect.com/science/article/pii/S1096751610000886)

Northcote, M. 2008. Online assessment: friend, foe or fix? *British Journal of Educational Technology, 33*(5):623 – 625. https://doi.org/10.1111/1467-8535.00335

Okada, A., Whitelock, D., Holmes, W., & Edwards, C. (2019). e-Authentication for online assessment: A mixed-method study. *British Journal of Educational Technology*, *50*(2), 861-875. https://doi.org/10.1111/bjet.12608

Padayachee, P., Wagner-Welsh, S., & Johannes, H. (2019). Online assessment in Moodle: A framework for supporting our students. *South African Journal of Higher Education*, *32*(5), 211-235. https://doi.org/10.20853/32-5-2599

Rahim, A.F.A. (2020). Guidelines for online assessment in emergency remote teaching during the COVID-19 pandemic. *Education in Medicine Journal. 12*(2):59–68. https://doi.org/10.21315/eimj2020.12.2.6

Sánchez-Mendiola et al., (2021). Evaluación del y para el aprendizaje a distancia: Recomendaciones para docentes de educación media superior y superior. CUAIEED, UNAM. Disponible en https://cuaieed.unam.mx/descargas/investigacion/evaluacion-del-y-para-el-aprendizaje-V02.pdf

Singh, V. & Thurman, A. (2019) How Many Ways Can We Define Online Learning? A Systematic Literature Review of Definitions of Online Learning (1988-2018), *American Journal of Distance Education*, *33*(4), 289-306. https://doi.org/10.1080/08923647.2019.1663082

The International Test Commission (2006) International Guidelines on Computer-Based and Internet-Delivered Testing, *International Journal of Testing*, *6*(2), 143-171. https://doi.org/10.1207/s15327574ijt0602_4

UNESCO (2020). *COVID-19. A glance of national coping strategies on high-stakes examinations and assessments.* Paris: UNESCO. https://en.unesco.org/sites/default/files/unesco_review_of_high-stakes_exams_and_assessments_during_covid-19_en.pdf

UNESCO (2021). *Evaluación formativa: Una oportunidad para transformar la educación en tiempos de pandemia. Reflexión a partir de los resultados del estudio cualitativo sobre perspectivas docentes en torno a la evaluación formativa.* OREALC/UNESCO Santiago. https://unesdoc.unesco.org/ark:/48223/pf0000378045

Zhou, K. (2016) *Non-cognitive skills: definitions, measurement and malleability*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000245576

Capítulo 10

EVALUACIÓN PROGRAMÁTICA

Silvia Lizett Olivares Olivares, Irma Elisa Eraña Rojas

"La evaluación es el motor del aprendizaje, ya que de ella depende tanto qué y cómo se enseña, cómo el qué y el cómo se aprende."
NEUS SANMARTÍ, 2007

CONCEPTUALIZACIÓN

La evaluación programática implica la cuidadosa selección y combinación de una variedad de métodos y actividades de evaluación, que están significativamente integrados en el diseño educativo, con el fin de promover el aprendizaje, así como obtener "una imagen completa" del desempeño de los alumnos (Schut et al., 2020).

La evaluación programática se genera de forma longitudinal, progresiva y sistemática, integrando de forma intencionada una colección de valoraciones del estudiante en un mismo lugar, ya sea un mapa, documento o espacio multimedia, en el cual se muestran los resultados de desempeño del estudiante a lo largo del programa. En la evaluación programática la información sobre la competencia de los alumnos se recopila intencional y continuamente para contar con la mayor evidencia posible sobre la evolución de su aprendizaje. La evaluación programática busca la forma de construir una valoración significativa y holística que refiera una conclusión del avance de cada competencia de los estudiantes en lugar de calificar fragmentos independientes. Por tanto, tiene un enfoque sistémico que observa la imagen completa en lugar de solo algunos de los pixeles (Schuwirth y Van der Vleuten, 2019).

Esta visión holística programática de la evaluación está enfocada en mejorar la validez, reproducibilidad y documentación del desarrollo de las competencias porque monitorea el progreso del estudiante hacia estándares de rendimiento definidos (Bok et al., 2013). La combinación de diferentes instrumentos de evaluación fortalece el proceso y la interpretación de los resultados de desempeño, ofreciendo un panorama general de progreso de las competencias del alumno. El proceso de evaluación se convierte en oportunidades de aprendizaje (evaluación para el aprendizaje) al mismo tiempo que permite toma de decisiones académicas (evaluación del aprendizaje) (Schuwirth y Van der Vleuten, 2011).

Schuwirth y Van der Vleuten (2019) describen la evolución histórica de la evaluación a través de varias décadas desde los exámenes de conocimientos hasta llegar a la evaluación programática, convirtiendo las pruebas de aprendizaje en estrategias para el aprendizaje. Según los autores, entre finales de los 60s y los 70s, se desarrollan en psicología diferentes pruebas psicométricas para determinar rasgos de la personalidad, los cuales se consideran como oportunidades para desarrollar pruebas de conocimientos con análisis de confiabilidad y validez. Los exámenes corresponden a una evaluación basada en ciencia en donde se privilegia el contenido. Para la década de los 90s, el Aprendizaje Basado en Competencias había cobrado relevancia con la necesidad de evaluaciones que integren conocimientos, habilidades y actitudes, sobre las cuales requiere observar comportamientos. Según Olivares Olivares y Valdez García (2017), hay un desacoplamiento entre la declaratoria del perfil de egreso y la formación en las aulas, con una cantidad de competencias excedidas en cuanto a las expectativas del plan de estudios. Dada la dificultad de evaluar competencias, las Actividades Profesionales Confiables *(Entrustable Professional Activities, EPAs* en inglés) permiten identificar tareas completas sobre las cuales se facilita la evaluación de competencias en un contexto de conceptualización profesional. En este sentido la evaluación programática facilita la identificación de instrumentos y alternativas para propósitos mixtos.

La evaluación programática tiene cuatro componentes esenciales:

- Evaluación por competencias en la evaluación programática.
- Seguimiento longitudinal vs. organización de la progresión curricular.
- Retroalimentación.
- Actividades de aprendizaje evaluables.

La evidencia mostrada sobre la efectividad de este tipo de evaluación y su impacto en el aprendizaje, según Van der Vleuten y Heeneman (2016) es que los estudiantes se convierten en buscadores activos de la retroalimentación y autogestionan su aprendizaje. Algunos de los principios que rigen este tipo de evaluación se muestran en la Tabla 1 de acuerdo con un consenso de expertos en la metodología (Heeneman et al., 2021).

Tabla 1. Principios de la evaluación programática

Principios	Descripción
Son múltiples puntos para la toma de decisiones.	Todas las evaluaciones son datos que forman parte de la evaluación programática, se utilizan para retroalimentación, y se agrupan para la toma de decisiones.
Metodologías mixtas de evaluación.	Los métodos son múltiples y su selección depende de su propósito.
Evaluación longitudinal.	El desempeño se registra de forma longitudinal sin distinguir si es formativa o sumativa.
Decisiones de progreso.	El rigor de la evaluación y la cantidad de puntos es proporcional a la decisión sobre la continuidad del progreso del estudiante. Se recomiendan comités para decisiones de filtro de avance o graduación.
Acompañamiento.	Los alumnos deben ser informados en todo momento de sus resultados y recibir orientación y asesoría sobre su desarrollo con mentores o asesores.

EVALUACIÓN POR COMPETENCIAS EN LA EVALUACIÓN PROGRAMÁTICA

La evaluación programática surge de la necesidad de obtener una visión completa de las competencias desarrolladas por el alumno y de optimizar la calidad de la evaluación. Una competencia es la integración de conocimientos, habilidades y actitudes para formar en contenidos disciplinares, transversales, interprofesionales o transdisciplinares (Olivares Olivares, 2021).

Las *competencias disciplinares* de los programas son las que imprimen el sello de la carrera, al establecer su carácter e identidad profesional. La identidad profesional es la evolución dinámica a través de la cual el alumno se reconoce a sí mismo como persona que forma parte de grupos sociales y académicos para ejercer su rol dentro de contextos profesionales y de esta forma generar un cambio consciente. En otras palabras, las competencias disciplinares son las que promueven los consejos u organismos acreditadores de los programas para formar al médico como médico, o al abogado como tal.

El concepto surge para la educación en salud, al considerar que las *competencias interprofesionales* son las necesarias para que dos o más profesiones aprenden entre sí, para mejorar la colaboración y el cuidado de la salud y de esta forma lograr una atención centrada en el paciente (WHO, 2010). Estas competencias son las que incluyen comunicación, trabajo en equipo y la construcción de acuerdos comunes para innovar y construir soluciones que beneficien a los usuarios.

Las *competencias transversales* son aquéllas que también se denominan genéricas o transferibles para permitirle al egresado desempeñarse en su contexto y a lo largo de la vida (Olivares Olivares et al., 2021a). Las *competencias transdisciplinares* son aquéllas que

permiten al egresado enfrentar los retos de la sociedad, lo que va más allá de las disciplinas individuales, para interiorizar el aprendizaje sobre el mundo real.

Para evaluar competencias bajo el enfoque programático, se requiere una selección cuidadosa de diversos métodos y un diseño sistémico organizacional (Schuwirth y Van der Vleuten, 2011). Las diferentes fuentes utilizadas proveen información sobre las fortalezas y debilidades de cada estudiante. Además de proveer evidencia del desarrollo de competencias profesionales sostenidas con un énfasis en resultados y enfocadas a un aprendizaje centrado en el estudiante.

La evaluación programática incluye los cuatro tipos de competencias, por lo tanto, va más allá de la toma de decisiones sobre únicamente competencias disciplinares. Considera la evolución en el grado de complejidad, y requiere interacciones con otros individuos en entornos de alta autenticidad para demostrar conocimientos, habilidades y actitudes que se van retroalimentando a medida que existe mayor práctica y retroalimentación.

Tradicionalmente se han evaluado las competencias en constructos separados tales como los conocimientos, habilidades y actitudes o profesionalismo, como si estos fueran estables y genéricos, sin contemplar la variabilidad que existe en las acciones de un individuo. Ser competente en un dominio es el resultado haber pasado a través de varios peldaños o módulos, cada uno con su respectiva evaluación.

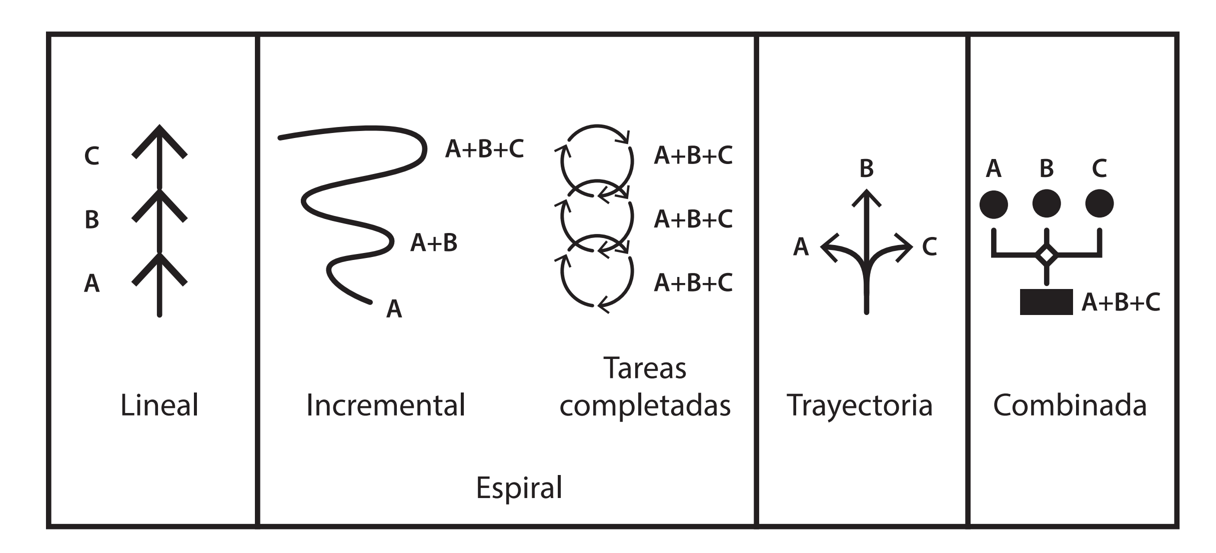
La mayor parte de los instrumentos actuales tienen limitaciones en cuanto a su reproducibilidad y consecuente confiabilidad en la interpretación de los resultados, por lo que su validez es limitada. Una sola medida es insuficiente, por tal motivo la combinación de instrumentos es ideal para que las múltiples fuentes den un panorama más holístico de las competencias desarrolladas. El trivializar el aprendizaje en busca de la objetividad enfocándose más en el formato más que en los contenidos ha llevado a perder la riqueza de la evaluación basada en un buen juicio profesional. Un muestreo optimizado y una combinación de información de distintas fuentes donde se involucre el juicio de evaluadores expertos que provean información cualitativa significativa. Este tipo de evaluación requiere profesores debidamente preparados.

La evaluación requiere estar alineada al currículo y a los objetivos del proceso de aprendizaje. Para esto se requiere un mapeo de las competencias a desarrollar y los momentos de aprendizaje que serán fuentes de información sobre el desempeño del alumno a lo largo de su plan de estudios. Los instrumentos de evaluación elegidos como fuentes de información deben ser debidamente seleccionados y estar alineados al objetivo de aprendizaje a medir.

SEGUIMIENTO LONGITUDINAL VS. ORGANIZACIÓN DE LA PROGRESIÓN CURRICULAR

En México, el Capítulo IV de las opciones educativas en educación superior, indica que "la trayectoria curricular es el orden y la manera en que deben lograrse los aprendizajes definidos institucionalmente para cada plan de estudio" (Granados Roldán, 2018). Para Olivares Olivares et al. (2021) y Olivares Olivares (2021) la organización de la progresión curricular puede clasificarse en lineal, espiral, trayectoria o una combinación de las anteriores, como se muestra en la Figura 1.

Figura 1. Organización del programa



Lineal

El modelo lineal es el más tradicional. Se centra en avanzar en contenidos secuenciales en donde cada curso disciplinar va sumando conocimiento al estudiante. Este modelo se considera como una progresión preestablecida o rígida, en donde se contempla una serie de asignaturas, módulos, o unidades de aprendizaje, y una secuencia única para cursarlas y acreditarlas (Granados Roldán, 2018).

En este tipo de progresión, aunque se enseñan de forma aislada, se espera que eventualmente los alumnos integren estos aprendizajes en las tareas de su práctica profesional. Como su nombre lo indica, la línea progresiva tiene un punto de inicio y uno de fin sin retorno, con lo cual es posible establecer un perfil de egreso acotado y fijo en el cual el control formativo puede mediarse para cumplir con ciertas competencias específicas establecidas por los organismos reguladores, asociaciones de afiliación, acreditadoras o consejos del programa o la profesión.

La forma más natural de acoplar la evaluación programática en un modelo curricular lineal es considerando un flujo de aprendizajes con entradas y salidas en serie, en donde se articulan diferentes momentos e instrumentos de evaluación según corresponda. El resultado es una lista de verificación generalmente dicotómica donde todas las competencias de egreso de un listado exhaustivo han sido evaluadas al menos una vez durante el programa. Es decir, si se tuviera que evaluar las competencias A, B y C, cada una de las mismas, tendría un resultado, A-Cumple, B-Cumple y C-Cumple. La barrera más importante para incorporar la evaluación programática en currículos lineales es pretender evaluar competencias que incluyen conocimientos, habilidades y actitudes cuando los contenidos analíticos solo consideran parcialidades de estas. Por tanto, la evaluación podría incorporarse como un recorrido de pequeñas partes de un proceso, en lugar de una integración de tareas o actividades profesionales completas.

Se recomienda utilizar aproximaciones menos rígidas para incorporar la evaluación programática en los programas educativos. La evaluación programática se adapta más a las teorías modernas de educación, como el aprendizaje constructivista y el desarrollo de competencias longitudinales con énfasis en la retroalimentación para optimizar el aprendizaje y la remediación para tender a las necesidades individuales de cada estudiante (Van der Vleuten et al., 2015).

Espiral

El modelo en espiral incrementa la complejidad en cada uno de los resortes que abre a medida que avanza el estudiante en el programa. Conserva el principio de puntas de una línea, una entrada y una salida, sin embargo, el espiral regresa a contenidos previos con lo cual se puede observar la progresión de las competencias bajo dos posibles opciones: a) incremental o b) tareas completas. El currículum en espiral lo describen Harden y Stamper (1999) como una forma iterativa de revisitar contenidos previos a través de un programa, buscando profundizar en ellos para construir un nuevo aprendizaje sobre los mismos. En este sentido, según los autores, la competencia progresa en cada uno de los momentos de evaluación.

Incremental, Espiral Cónica

La estructura incremental desarrolla la evolución de los aprendizajes en secuencia que va llevando al alumnado de una etapa de novato a competente acumulando contenidos previos. El nivel de dominio de la competencia se va desarrollando gradualmente, de la misma manera, la evaluación va midiendo su desarrollo. Cada momento de evaluación individual sirve como fuente de información que se van combinando para evaluar la progresión del alumno. El desarrollo de dichas competencias es monitoreado por un mentor o asesor. Esta forma de evaluar el aprendizaje incluye pensar en evaluar la competencia A, después A+B y al final A+B+C, integrando una competencia a la anterior, en donde el espiral se abre en forma de figura cónica expandiendo su diámetro, retomando aprendizajes previos, pero ampliando los consecutivos.

Integración en Perspectivas o Tareas, Espiral Cilíndrica.

La intención de la evaluación programática es que sea longitudinal, y por tanto las evaluaciones se integran en conductas observables que mezclan teórica-práctica en múltiples momentos del programa. Algunos conceptos que se utilizan para esta integración son: Actividades Profesionales Confiables, Aprendizaje Basado en Perspectivas, Tareas completas o Exámenes Progresivos. La característica común es que el espiral guarda la forma de un cilindro en el cual se miden los mismos constructos en cada momento de evaluación: A+B+C, A+B+C, A+B+C.

Las EPAs son la forma de convertir competencias en actividades observables y medibles en la práctica clínica (Shorey et al., 2019). Al pasar de modelos curriculares lineales a modelos integrados como el de espiral, las EPAs facilitan la evaluación progresiva porque permiten valorar la evolución del estudiante en cierta actividad hasta que se consideran

que puede desarrollarla sin la necesidad de supervisión y por tanto se le puede "confiar" su ejecución, sin riesgos para el paciente. Dado que las competencias son abstractas y tienen diferente énfasis en lo conceptual, procedimental o actitudinal, las EPAs se introducen como mecanismo de valoración porque son observables, medibles y cuentan con un contexto de autenticidad comprensible para todos los que intervienen en el proceso de evaluación.

El Aprendizaje Centrado en el Paciente atiende al concepto de perspectivas porque se evalúan múltiples competencias en cada encuentro con el paciente. El Aprendizaje Centrado en el Paciente es un modelo formativo que busca entender al paciente como persona considerando su individualidad, estructura de significados y contexto desde las perspectivas: humana, biomédica y de gestión para atender sus necesidades de forma integral y generar el emprendimiento de cambios innovadores que favorezcan su realidad y la de aquellos en condiciones equivalentes (Olivares Olivares y Valdez García, 2017). En este sentido en cada encuentro se evalúa en forma simultánea las cuatro perspectivas. La humana, la cual considera aspectos de empatía, comunicación y trato considerando al paciente como un individuo indivisible en cuando a lo físico y mental. La biomédica, la cual considera la forma de comprender la mejor evidencia científica relacionada con el motivo de consulta del paciente. La de gestión la cual administra cada uno de los recursos y regulaciones para actuar de forma ética, eficiente y segura. Y la de emprendimiento, perspectiva que permite potenciar y mejorar las alternativas para prevenir y mejorar la salud de todos aquellos que enfrenten una condición equivalente. En este sentido, se evalúan las cuatro perspectivas de forma integrada en cada encuentro clínico virtual, simulado o real.

El concepto de tareas completas (*Whole-Tasks*) busca integrar actividades de aprendizaje para eliminar la fragmentación de contenidos disciplinares. El modelo 4C/ID (*4 Components / Instructional Design*) incluye cuatro componentes: a) Tareas de aprendizaje, b) Práctica de partes de una tarea, c) Información de apoyo y d) Información procedimental. Las tareas de aprendizaje son la integración de competencias recurrentes (conocimientos, habilidades y actitudes), basadas en experiencias auténticas que se organizan de simples a complejas. Antes de avanzar a un nuevo nivel de complejidad, las tareas varían en su estímulo reduciendo la supervisión del docente hacia el estudiante hasta que es capaz de realizarla en forma autónoma. Las partes de una tarea proveen prácticas psicomotoras para ejercitar tareas estrictamente manuales. La información de soporte y la información procedimental se refieren a los contenidos teóricos y prácticos respectivamente, que requieren los estudiantes para el aprendizaje de la tarea.

El concepto de exámenes progresivos (*progress testing*) atienden a una integración de tipo espiral cilíndrico. Es lo equivalente a presentar el examen final antes de llevar los contenidos en cualquiera los momentos de evaluación. Los exámenes progresivos se introdujeron en la Universidad de Maastricht para evaluar el aprendizaje de los estudiantes con el Aprendizaje Basado en Problemas (Freeman et al., 2010). De acuerdo con los autores, esta modalidad impulsa a los alumnos a mayor dedicación en el estudio, así como a generar una comparación con otras universidades. Si bien este modelo de evaluación puede

provocar críticas por incluir reactivos de temáticas que los alumnos no reconocen, también involucra ítems de contenidos que se aprendieron en el pasado, en este sentido se puede valorar el deterioro o retención de aprendizajes, así como múltiples variables de investigación educativa tales como el rigor de las pruebas, comparaciones de desempeño entre individuos, generaciones, métodos de enseñanza, instrucción docente, medios de entrega, etcétera.

TRAYECTORIAS FLEXIBLES

Un modelo de trayectorias busca flexibilizar las oportunidades de aprendizaje en el programa. En lugar de tener predefinidos cada uno de los cursos dentro del programa, se cuenta con un menú de oportunidades de aprendizaje que pueden ser seleccionados según las preferencias del estudiante. Según el acuerdo 18/11/18 una trayectoria flexible es en la que el estudiante elige las asignaturas, módulos, o unidades de aprendizaje a cursar y el orden en el que las acredita (Granados Roldán, 2018). Utilizar trayectorias flexibles se puede asociar a una comparación a pasar de un álbum a una lista de reproducción tipo *playlist* (Olivares Olivares, 2021). El álbum es lo equivalente a un programa lineal en donde tiene que pasar por todas las canciones en un orden predeterminado. En una lista de reproducción (*playlist)* las canciones se mezclan a gusto del usuario y solo se incluyen aquellas que se desea incluir. A diferencia de los programas lineales o en espiral, donde las competencias de egreso pueden ser predefinidas; en un modelo abierto o flexible resulta imposible predecir cuáles son las competencias de egreso, dado que depende de los cursos que terminen siendo seleccionados por el estudiante. En este sentido, la evaluación programática, tampoco puede ser preestablecida dado que es un mosaico de tantas combinaciones como la oferta educativa lo establezca. Para resolverlo, se necesita utilizar unidades de aprendizaje que no tienen requisito o procedencia de cursos previos y las evaluaciones se convierten en coleccionar logros tipo insignias durante los estudios. La flexibilidad debe considerar espacios propicios para la remediación.

COMBINADA

Según el acuerdo 18/11/18, la trayectoria combinada es cuando existe un determinado número de asignaturas, módulos, o unidades de aprendizaje que se consideran en orden de acreditación seriado y, en que el resto, el estudiante elige el orden en el que las cursa y acredita (Granados Roldán, 2018). Las tendencias curriculares implican que las opciones flexibles y combinadas son las que estarán imperando de forma más común en el futuro. Para incorporar la evaluación programática en un modelo combinado es relevante utilizar mezcla de instrumentos, que consideren espirales integrados como insignias sueltas a lo largo del programa, dependiendo de la enseñanza otorgada. Esto implica que se requiere alineación tanto en el marco de competencias y el currículo, balanceando los elementos tradicionales modulares y los longitudinales.

Para un cambio hacia una evaluación programática de cualquier tipo de trayectoria se requiere involucrar a todos los agentes relacionados con la evaluación, revisión de normativas y regulaciones. En este sentido, Van der Vleuten y Heeneman (2016) proponen como una alternativa una implementación parcial, estableciendo la evaluación programática en el ambiente clínico y no en las aulas. Otras implementaciones parciales incluyen el agregar más retroalimentación a programas de evaluación o introducir un monitoreo de estudiantes o un sistema de mentores.

RETROALIMENTACIÓN

En la evaluación programática se promueve buscar evaluaciones que fomenten la orientación y la retroalimentación sobre la calidad de su aprendizaje. Cada momento de evaluación en la evaluación programática es una oportunidad para dar retroalimentación de alta calidad, es un componente crucial de este tipo de evaluación (Van der Vleuten et al., 2015).

El seguimiento del alumno se da mediante un mentor o un asesor que revisa el progreso del alumno a través del programa. El mentor puede guiar al alumno en una reflexión sobre su progresión en el plan de estudios y en conjunto planear las nuevas metas de aprendizaje (Dornan, 2014). Parte de las actividades del mentor incluyen el monitoreo de la progresión, el diálogo reflexivo, la planeación de actividades de remediación y el seguimiento subsecuente.

Existen distintas formas en las que se puede dar retroalimentación significativa de la progresión, en exámenes estandarizados se utiliza el reporte con información detallada de las categorías de temas y subtemas o de los objetivos de aprendizaje. En el ECOE (Examen Clínico Objetivo Estructurado) a través de los dominios de habilidades o a través de retroalimentación oral o escrita. Para evaluaciones de habilidades complejas como la comunicación y el profesionalismo se puede dar retroalimentación narrativa con el uso de preguntas abiertas.

El mentor y el alumno no son los únicos actores, en la retroalimentación pueden involucrarse reflexiones entre pares o en pequeños grupos. La reflexión entre en comunidad proveen un espacio de retroalimentación y aprendizaje. Así mismo esto puede implementarse en pequeños grupos siempre y cuando se establezca un ambiente de acompañamiento y apoyo en su aprendizaje.

Schuwirth y Van der Vleuten (2011) presentan la analogía de las decisiones del médico clínico. El médico utiliza más de un estudio para diagnosticar a su paciente, sobre los cuales no considera una suma o un promedio de los resultados de laboratorio para decidir si el paciente está saludable o no. En contraste, analiza en conjunto la información relevante de la historia clínica, examen físico y resultados de laboratorio para determinar los pasos siguientes para la intervención y tratamiento. Esta aproximación de considerar múltiples fuentes de información es uno de los principios para la evaluación programática.

ACTIVIDADES DE APRENDIZAJE EVALUABLES

En la evaluación programática se obtiene información de las actividades de aprendizaje desarrolladas por el alumno en diferentes entornos de aprendizaje, a través de diferentes métodos de evaluación y con diferentes evaluadores. Se enfatiza la importancia de estructurar todos los procesos de evaluación e instrumentos de tal manera que al agregarlos provean información significativa ya sea cualitativa o cuantitativa para ser posteriormente juzgada por un comité.

Algunos ejemplos de cómo se podría utilizar la evaluación programática con la finalidad de concluir si un alumno ha progresado en cierto dominio de habilidades es a través de los resultados de exámenes de opción múltiple, evaluaciones de simulación o el mini-CEX (observación en el contexto clínico). Mediante este tipo de evaluaciones se pueden abordar, tanto los niveles logrados de los dominios evaluados como aquellos que están en proceso de desarrollo.

Durante las rotaciones clínicas cada momento de aprendizaje es una oportunidad para llevar a cabo una evaluación programática. Esta congruencia de objetivos entre evaluación y actividades le da mayor significado al aprendizaje y apoya el concepto de evaluación para el aprendizaje (Van der Vleuten et al., 2015). Por ejemplo, Lucio et al. (2020) presentan una aplicación en línea para evaluar a cada estudiante en tiempo real durante sus rotaciones clínicas. Su proyecto incluye múltiples jueces para valorar el desempeño a partir del Aprendizaje Centrado en el Paciente, lo cual considera cuatro perspectivas: biomédica, humana, gestión y emprendimiento, en múltiples momentos clínicos tales como entrega de guardia, intervenciones quirúrgicas, emergencias, pase de visita y consulta. Como resultado, es posible determinar el desempeño longitudinal del estudiante, el patrón de evaluaciones del profesor y los escenarios más significativos para el aprendizaje. El agrupamiento de la información resultante se almacena en portafolios electrónicos que pueden servir como repositorio formal e informal para evaluación y retroalimentación, para facilitar aspectos administrativos y logísticos del proceso de evaluación y para ser un visor práctico de la información almacenada.

Las múltiples fuentes de información provista y combinada requieren posteriormente un juicio de un profesional experto. En este sentido representan la evidencia para la toma de decisiones, similar a la práctica actual en la atención médica de medicina basada en evidencia (Schuwirth et al., 2017). Dado a que las decisiones que se toman con dichos resultados son de alta relevancia para la progresión del alumno, se requiere que dicho juicio sea creíble y confiable. Por lo que se sugiere que las decisiones se tomen en un comité de evaluación, limitando la participación del mentor en la toma de decisión final, para prevenir conflictos de interés.

Las discusiones dentro del comité toman mayor importancia en particular en los casos donde una decisión se tenga que tomar entre aprobar y reprobar, es decir, esta información se considera como filtro de progreso de un estudiante para continuar en el programa o solicitar su baja. Al contar con información actualizada y de múltiples fuentes, en caso de tomarse una decisión no aprobatoria, esta no debiera ser una sorpresa para el alumno ni para el mentor, ya que se cuenta con retroalimentación constante y sesiones informativas intermedias

entre ambos. Este sistema es tan robusto y con buena credibilidad que, a pesar de crear un sistema de apelaciones, estas raras veces suceden (Van der Vleuten y Heeneman, 2016).

Los resultados de las evaluaciones programáticas también resultan de utilidad para los procesos de cartas de recomendación a egresados, así como procesos de regulación, acreditación o premiación de los programas académicos.

RETOS DE LA EVALUACIÓN PROGRAMÁTICA

La evaluación programática no es sencilla de implementar, requiere que los diferentes grupos de interés logren captar los fundamentos detrás de la misma. La evaluación programática se puede utilizar tanto en estudios de pregrado, posgrado y educación profesional continua en medicina y en otros campos de las ciencias de la salud (Schuwirth y Van der Vleuten, 2011).

El costo es otra de las dificultades, al requerir de mentores, retroalimentación individual y decisiones mediante un comité de expertos, se requieren una buena administración de recursos. Entre estos, el tiempo que le dedican todos los involucrados, el trabajo para su implementación es significativo, además del entrenamiento de los mentores y evaluadores, la creación y llenado del portafolio y la recolección de información.

Para asegurar la calidad de la evaluación ya no basta solo con los criterios de reproducibilidad y validez, se debe incluir equidad, confiabilidad y confianza determinados a través de procesos organizacionales tales como segundas opiniones, observaciones independientes o consultas interdisciplinarias (Driessen et al., 2005). Otros criterios que se deben de observar en toda evaluación son el costo-beneficio, impacto educativo y la aceptabilidad, salvaguardando la confiabilidad y credibilidad de la decisión sobre el desempeño de un alumno (Norcini et al., 2018).

Por último, para el seguimiento de cada estudiante es necesario utilizar habilitadores que apoyen la implementación y visibilidad de los resultados obtenidos en la trayectoria de cada estudiante y contar con la posibilidad de estratificarlos de múltiples formas, por año, temática, competencias, desempeño, etc.

EVALUACIÓN PROGRAMÁTICA: CASO DE LA ESCUELA DE MEDICINA Y CIENCIAS DE LA SALUD, TECNOLÓGICO DE MONTERREY

La Escuela de Medicina y Ciencias de la Salud en sus planes 2019 cuenta con una evaluación programática que incluye:

- Evaluación de competencias en unidades de formación.
- Evaluación de competencias en cierre de etapa.
- Examen Integrador Incremental Semestral (EIIS).

El Modelo Educativo Tec21 de los programas de salud del Tecnológico de Monterrey cuentan con un modelo de evaluación por competencias que se estratifican en disciplinares (por

programa), interprofesionales (o de área) y transversales (formación de la persona y transdisciplinares). Las competencias interprofesionales son cuatro (Olivares Olivares, 2021):

1) **Investigación en salud**: Aplica conocimiento actualizado y el método científico para generar propuestas innovadoras a los problemas prevalentes de salud, con pensamiento crítico y perspectiva multidisciplinaria.
2) **Atención Centrada en el Paciente**: Aborda integralmente a personas y grupos en la detección temprana de padecimientos, factores de riesgo o situaciones de primeros auxilios, generando acciones de cambio que favorezcan estilos de vida saludable.
3) **Profesionalismo:** Se desempeña con profesionalismo en la interacción con pacientes, familias y equipos multidisciplinarios, en entornos comunitarios y asistenciales, locales y globales, reconociendo cada uno de los roles de colaboración.
4) **Gestión en salud:** Gestiona los procesos y recursos requeridos en las diferentes fases de la atención clínica con eficiencia, calidad y seguridad para el grupo de interés y el personal de salud.

Cada una de estas competencias se evalúan en unidades de formación a través de los retos. El Aprendizaje Basado en Retos es el cual es un modelo formativo donde los alumnos colaboran activamente para resolver problemas actuales que aquejan a la sociedad, a partir de la indagación e integración de contenidos teórico-prácticos que se transforman en propuestas creativas e innovadoras que impactan el público seleccionado (Olivares Olivares, 2021). Para logarlo, los retos cuentan con la participación entre docentes, estudiantes y socios formadores. Los socios formadores son individuos, grupos, comunidades y organizaciones que proporcionan los retos a enfrentar y pueden tomar un papel activo como usuarios, facilitadores expertos y/o jueces de la evaluación.

El Modelo Educativo Tec21 tiene un diseño curricular en trayectorias. La evaluación de etapa sucede en tres momentos para cada estudiante y coincide con la última semana del semestre (Semana 18) en que termina una fase consecutiva: Exploración, Enfoque y Especialización. El proceso de implementación de la *Semana 18 Cierre de Etapa de Exploración* considera la utilización de evaluación auténtica para que el estudiante demuestre lo aprendido a partir de *situaciones de la vida real*, *problemas significativos* y *planteamientos complejos del entorno* (Olivares Olivares et al., 2021b). Algunos ejemplos son portafolio digital de evidencias, escenario situacional y entrevista por competencias. El propósito de la formación del Cierre de Etapa es formativo e incluye una retroalimentación individual a cada estudiante y una reflexión sobre sus logros y áreas de oportunidad.

En forma adicional, además de los bloques con la orientación de Aprendizaje Basado en Retos se cuenta con el Examen Integrador Incremental Semestral (EIIS) que es un instrumento de evaluación de las competencias cognitivas de los estudiantes de salud en el Modelo Educativo Tec21. Utiliza los principios de la evaluación progresiva y de mejoramiento de pruebas (*Test Enhacement Learning*). Se aplica tres veces al semestre incluyendo el contenido

total de todas las materias que cursan en el periodo correspondiente. El EIIS se diseña por la comunidad académica de profesores que imparten las materias a nivel nacional quienes definen y validan cada uno de los reactivos. Los contenidos se asocian con los objetivos de aprendizaje de cada materia, según el nivel de avance y complejidad en el plan de estudios. Los reactivos se estratifican para que haya un balance entre temas y grado de dificultad. Para que el EIIS tenga calidad y consistencia, utiliza un grado de discriminación de conocimiento, es decir, que tenga la capacidad de medir los aprendizajes que desea medir de acuerdo con los objetivos del curso.

CONCLUSIONES Y REFLEXIÓN

La evaluación programática optimiza tanto la toma de decisiones sobre el desempeño de los alumnos como la función de aprendizaje de la evaluación. Enriquece los argumentos que nutren la evaluación robusteciendo así el sistema para la toma de decisiones sobre la progresión del aprendizaje en cada alumno.

La evaluación programática tiene la ventaja de otorgar a los alumnos múltiples oportunidades formativas para reflexionar sobre sus fortalezas y áreas de oportunidad. Sin embargo, para que el proceso sea exitoso es necesario establecer, por diseño curricular y de los habilitadores, las especificaciones concretas con las que se concibe dicha evaluación para forme parte del proceso de enseñanza-aprendizaje y no se considere como sobrecarga o actividades adicionales a las proyectadas en el programa.

REFERENCIAS

Bok, H. G., Teunissen, P. W., Favier, R. P., Rietbroek, N. J., Theyse, L. F., Brommer, H., Jaarsma, D. A. (2013). Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Medical Education, 13*, 123-132. https://bmcmededuc.biomedcentral.com/articles/10.1186/1472-6920-13-123

Dornan, T. M. (2014). *Medical Education: Theory and Practice: Theory and Practice.* UK: Elsevier Health Sciences.

Driessen, E., Van der Vleuten, C., Schuwirth, L. V., & Vermunt, J. (2005). The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: A case study. *Medical Education, 39*(2), 214–220. doi: https://doi.org/10.1111/j.1365-2929.2004.02059.x

Freeman, A., Van der Vleuten, C., Nouns, Z., & Ricketts, C. (2010). Progress testing internationally. *Medical Teacher, 32*(451-455), 451-455.

Granados Roldán, O. R. (2018). *Acuerdo número 18/11/18 por el que se emiten los lineamientos por los que se conceptualizan y definen los niveles, modalidades y opciones educativas del tipo superior.* Ciudad de México: Secretaría de Educación Pública.

Harden, R., & Stamper, N. (1999). What is a spiral curriculum? *Medical Teacher, 21*(2), 141-143.

Lucio, C. A., Nigenda, J. P., Garcia-Garcia, M., & Olivares, S. L. (2020). Clinical competence assessment: development of a mobile app to enhance patient centerderness. *Development and Learning in Organizations, 34*(2), 17-20. https://doi.org/10.1108/DLO-08-2019-0186

Norcini, J., Anderson, M. B., Bollela, V., Burch, V., Costa, M. J., Duvivier, R., Swanson, D. (2018). Consensus framework for good assessment. *Medical Teacher, 40*(11), 1102–1109. https://doi.org/10.1080/0142159X.2018.1500016

Olivares Olivares, S. L. (2021). *Aprendizaje Basado en Retos: Transformando la educación en salud.* McGraw Hill. Obtenido de https://www.mheducation.com.mx/vs-aprendizaje-basado-en-retos-transformar-la-educacion-en-salud-9786071515940-latam

Olivares Olivares, S. L., y Valdez García, J. E. (2017). *Aprendizaje Centrado en el Paciente: Cuatro perspectivas para un abordaje integral.* Editorial Médica Panamericana.

Olivares Olivares, S. L., López Islas, J. R., Pineda Garín, M. J., Rodríguez Chapa, J., Aguayo Hernández, C., y Peña Ortega, L. (2021a). *Modelo Educativo Tec21: Retos para una vivencia que transforma.* Monterrey, México: Editorial Digital Tecnológico de Monterrey.

Olivares Olivares, S. L., Zubieta Ramírez, C., Huereca Alonzo, S., Mendoza Moreno, J. A., y Aguayo Hernández, C. H. (2021b). Cerrando la Etapa de Exploración: Desarrollo de la identidad profesional Autoconocimiento y relaciones afines. *Congreso Latinoamericano de Educación ELE.* Cartagena, Colombia: ELE.

Schut, S., Maggio, L. A., Heeneman, S., Tartwijk, J., & Van der Vleuten, C. (2020). Where the rubber meets the road: An integrative review of programmatic assessment in health care professions education. *Perspectives in Medical Education*. https://doi.org/10.1007/s40037-020-00625-w

Schuwirth, L. W., & Van der Vleuten, C. P. (2011). Programmatic assessment: From assessment of learning to assessment for learning. *Medical Teacher, 33*(6), 478–485. https://doi.org/10.3109/0142159X.2011.565828

Schuwirth, L., & Van der Vleuten, C. (2019). How 'Testing' Has Become 'Programmatic Assessment for Learning'. *Health Professions Education, 5*, 177–184.

Schuwirth, L., Van der Vleuten, C., & Durning, S. J. (2017). What programmatic assessment in medical education can learn from healthcare. *Perspectives in Medical Education, 6*, 211–215. https://doi.org/10.1007/s40037-017-0345-1

Shorey, S., Lau, T., Lau, S., & Ang, E. (2019). Entrustable professional activities in health care education: a scoping review. *Medical Education*, 766-777. https://doi.org/10.1111/medu.13879

Van der Vleuten, C. P., & Heeneman, S. (2016). A new holistic way of assessment: programmatic assessment. *FEM Fundación Educación Médica, 19*(6), 275-279.

Van der Vleuten, C., Schuwirth, L., Driessen, E., Govaerts, M., & Heeneman, S. (2015). Twelve tips for programmatic assessment. *Medical Teacher, 37*, 641–646.

WHO (2010). *Framework for action on Interprofessional Education & collaborative practice.* World Health Organization. Health Professions Networks Nursing & Midwifery Human Resources for Health.

Capítulo 11

EVALUACIÓN Y PENSAMIENTO DE SISTEMAS

Melchor Sánchez Mendiola, José Gerardo Moreno Salinas

"...el pensamiento sistémico es una disciplina para ver totalidades. Es un marco para ver interrelaciones en lugar de cosas, para ver patrones de cambio en lugar de instantáneas estáticas."

Peter Senge, "La Quinta Disciplina"

INTRODUCCIÓN

El profesorado cuenta con gran cantidad de instrumentos para evaluar y propiciar el aprendizaje de los estudiantes. Estas herramientas utilizan principios pedagógicos y técnicas diferentes, que a su vez requieren diversas metodologías y marcos conceptuales para su diseño, aplicación y utilización. El tema central de este capítulo es la evaluación y la importancia de desarrollar una perspectiva sistémica de la misma. Se revisará el concepto fundamental de sistema y pensamiento de sistemas, se describirán ejemplos y por medio de diferentes cuestionamientos, se reflexionará sobre la importancia de desarrollar un enfoque sistémico e integrador de la evaluación educativa.

Se utilizará el ejemplo de la evaluación de un curso masivo abierto en línea (MOOC, por sus siglas en inglés), en el que se pretende evaluar el aprendizaje de los participantes y el éxito del curso. Para ello deben hacerse las siguientes preguntas: ¿qué instrumentos usar?, ¿cómo articular los instrumentos de evaluación con el diseño educativo del curso?, ¿cómo integrar los resultados con todo el proceso de aprendizaje de individuos y grupos?, ¿cómo realimentar el curso y a los docentes que lo diseñaron para mejorar la calidad del mismo? Para dar respuesta a estas interrogantes se pueden realizar las siguientes acciones: documentar el número de participantes que visualizan todo el material, cuántos participantes terminan el curso y obtienen un certificado, aplicar exámenes de opción múltiple formativos y sumativos sobre los temas revisados, analizar cualitativamente las discusiones en foros, aplicar cuestionarios de seguimiento para verificar si aplican lo aprendido en la práctica, entre muchas otras posibilidades.

¿Cómo integrar toda la información de evaluación obtenida por los diversos instrumentos y en diferentes momentos del proceso educativo? Es fácil perderse en el bosque de la evaluación educativa, con tantos árboles, ramas y hojas de métodos, instrumentos y diferentes

perspectivas. Por ello es fundamental desarrollar una **perspectiva sistémica** del proceso educativo y de los elementos de enseñanza, aprendizaje y evaluación que lo integran, ya que existe una compleja interacción entre todos ellos (Aparicio et al., 2021; Bowe y Armstrong, 2017; Ndaruhutse et al., 2019). En el ejemplo descrito para evaluar un MOOC, un esquema tradicional sin perspectiva sistémica podría ser el siguiente: diseñar elementos de evaluación formativa y sumativa para los objetivos de aprendizaje de cada unidad temática, analizar los resultados de cada examen o aplicación de instrumento, acreditar cada unidad si se aprueban las evaluaciones sumativas, obtener el certificado si se aprueban todas las evaluaciones sumativas del curso. El esquema tradicional solamente sigue la secuencia de actividades programadas de evaluación, sin realizar esfuerzos explícitos por integrar toda la información en un todo coherente que funcione para realimentar el proceso y mejorarlo. El diseño de la evaluación refleja el modelo mental utilizado, por lo que, si solo se tiene un martillo, todo parece un clavo y se actúa en consecuencia (Figura 1).

Figura 1. Esquema de la visión tradicional de la evaluación en educación (adaptado de Bowe, 2019)

Una de las principales limitaciones de esta práctica tradicional de la evaluación, es que cada unidad temática del curso o elemento del currículo incluye de forma rutinaria un componente de evaluación, generalmente sumativo, que al final se suma con los demás para otorgar una calificación. Por ello es necesario ampliar la perspectiva de la evaluación, ya que si se continúa con estos usos y costumbres se pierde la oportunidad de aprovechar las herra-

mientas que proporciona la aproximación sistémica, que pueden incrementar la calidad del proceso, impactar en el aprendizaje de los estudiantes y mejorar las interacciones entre los elementos del sistema educativo.

LOS SISTEMAS, SU ENFOQUE Y EL PENSAMIENTO SISTÉMICO

El biólogo alemán Ludwig von Bertalanffy es considerado el padre de los sistemas, en 1968 desarrolló la Teoría General de los Sistemas (TGS), y cambió el paradigma de que los sistemas en su conjunto son iguales a la suma de sus partes (Aparicio et al., 2021). Dicha TGS ha sido aplicada tanto a áreas del conocimiento de las ciencias exactas como de las ciencias sociales, con el propósito de estudiar las interacciones de sus elementos constituyentes. Ackoff y Emery (1972), precisan que la TGS y sus interacciones se caracterizan por adoptar los siguientes conceptos: entradas (insumos), salidas (productos), procesos de transformación y retroalimentación. Las entradas también son consideradas como los consumibles (información, materiales o energía) que se requieren para producir resultados. Los procesos de transformación son los mecanismos o tecnologías utilizados para convertir las entradas en salidas. Las salidas son los resultados obtenidos por los componentes del sistema que conducen a su contribución al medio ambiente. Las salidas también pueden generar retroalimentación convirtiéndose en entradas del sistema.

Una definición comúnmente aceptada de sistema es el conjunto de elementos que están íntimamente interrelacionados y que tienen un propósito en común (Dawidowicz, 2012; Norman, 2013; Plack et al., 2018, 2019). Por ejemplo, tenemos el sistema planetario, el sistema nervioso, el sistema de educación, el sistema Universidad Abierta y Educación a Distancia (SUAyED) de la Universidad Nacional Autónoma de México (UNAM), etcétera. Una definición similar, utilizada en evaluación educativa, es: "un sistema es un conjunto de cosas diferentes que, funcionando en conjunto, producen un resultado que no se puede lograr por cada cosa por separado" (Bowe y Armstrong, 2017). Uno de los valores agregados más importantes del pensamiento sistémico es el resultado de las interacciones que ocurren, de forma dinámica, entre las fases y elementos del sistema.

No solo basta con reconocer los elementos (estructura) que los constituyen, sino cómo es que estos se interrelacionan (procesos) dentro de cada uno de ellos para cumplir con los propósitos para los cuales los sistemas fueron concebidos, ya sean por sus fenómenos físicos, composición química, biológica o por el ser humano. Independientemente de la disciplina, un sistema se caracteriza por (1) estar formado por un conjunto de elementos, (2) los elementos están relacionados entre sí, (3) operan en un determinado entorno, (4) tener objetivos comunes, y (5) capacidad de auto-organización (Aparicio et al., 2021).

Es fundamental tener claridad que el todo es más que la suma de sus partes, dicho en otras palabras, la integración de varios elementos da como resultado beneficios superiores a los que generarían la suma individual de cada uno de ellos por separado (Colbert et al., 2011; Norman, 2013). Es así que se puede tener una mayor conciencia del fenómeno o sistema de estudio. Posibilitando un enfoque sistémico, de acuerdo con Rosnay (1977) lo define

como el análisis de la interacción que ocurre entre elementos y el medio ambiente, con el supuesto que las relaciones son multidireccionales, no lineales y van más allá de una relación causa-efecto.

De acuerdo con la Asociación Médica Estadounidense (AMA por sus siglas en inglés), los pensadores sistémicos deben desarrollar los siguientes hábitos (AMA, 2019):

- Comprender el panorama general.
- Observar cómo los elementos del sistema cambian en el tiempo, generando patrones y tendencias.
- Reconocer que la estructura de un sistema genera comportamiento.
- Identificar la naturaleza circular de las relaciones complejas de causa y efecto.
- Establecer conexiones significativas dentro y entre sistemas.
- Cambiar de perspectivas para aumentar la comprensión.
- Supuestos superficiales y de prueba.
- Considerar el problema en su totalidad y resistirse a llegar a una conclusión rápida.
- Considerar cómo los modelos mentales afectan la realidad actual y el futuro.
- Utilizar la comprensión de la estructura del sistema para identificar posibles acciones de impulso o aprovechamiento.
- Considerar las consecuencias de las acciones a corto, largo plazo y no deseadas.
- Poner atención hacia las acumulaciones y sus tasas de cambio.
- Reconocer el impacto de los retrasos de tiempo al explorar relaciones de causa efecto.
- Revisar los resultados y cambios de las acciones de ser necesario, proceso conocido como "aproximación sucesiva".

El pensamiento de sistemas es comprender cómo diferentes componentes y elementos de un sistema interactúan e impactan unos con otros, va más allá de mapear a los participantes, ya que incluye analizar las conexiones formales e informales, y cómo estas impactan el funcionamiento del sistema (Ndaruhutse et al., 2019). Esta estrategia puede ser particularmente poderosa cuando un problema es importante, crónico, o ha sido tratado de resolver anteriormente sin éxito. Es diferente del pensamiento lineal tradicional, como se describe en la Tabla 1.

Tabla 1. Comparación del pensamiento lineal tradicional y del pensamiento de sistemas (adaptado de Ndaruhutse et al., 2019)

Pensamiento lineal tradicional	Pensamiento de sistemas
Ve las partes individuales del todo, con frecuencia de forma aislada.	Ve al sistema como un todo.
Se enfoca principalmente en el contenido.	Se enfoca principalmente en los procesos.
Adopta un enfoque de causa-efecto e intenta resolver los síntomas en lugar de los problemas subyacentes.	Busca entender las causas potenciales y los factores dinámicos que pueden estar en juego, incluyendo bucles de retroalimentación.
Tiende a pensar "técnicamente" sobre los problemas y creer que el problema es solucionable con un procedimiento sencillo.	Tiende a pensar "panorámicamente" (la "vista de pájaro") e incluye tomar en cuenta incentivos, aspectos políticos, económicos, sociales, al pensar qué tan difícil puede ser resolver un problema.

Desarrollar una perspectiva sistémica de la evaluación permite ver el panorama global del aprendizaje de forma amplia y profunda, y arroja el valor agregado de las interacciones en las interfaces de los diversos elementos del sistema, que generan información valiosa. A diferencia del pensamiento tradicional lineal, que simplifica en exceso las causas y efectos de las cosas, el pensamiento de sistemas abre una ventana a un mundo diferente, más auténtico, más sofisticado, con mayores elementos para interpretar la enseñanza y la evaluación del aprendizaje en su contexto complejo (Richmond, 1993; Senge, 2006; Stave y Hopper, 2007).

DISEÑO Y PROPÓSITOS FUNCIONALES DE LOS SISTEMAS EN EVALUACIÓN

De acuerdo con Bowe (2017, 2019), existen cinco elementos de diseño y propósitos funcionales de los sistemas que aplican en los escenarios educativos de evaluación, que se describen en la Tabla 2.

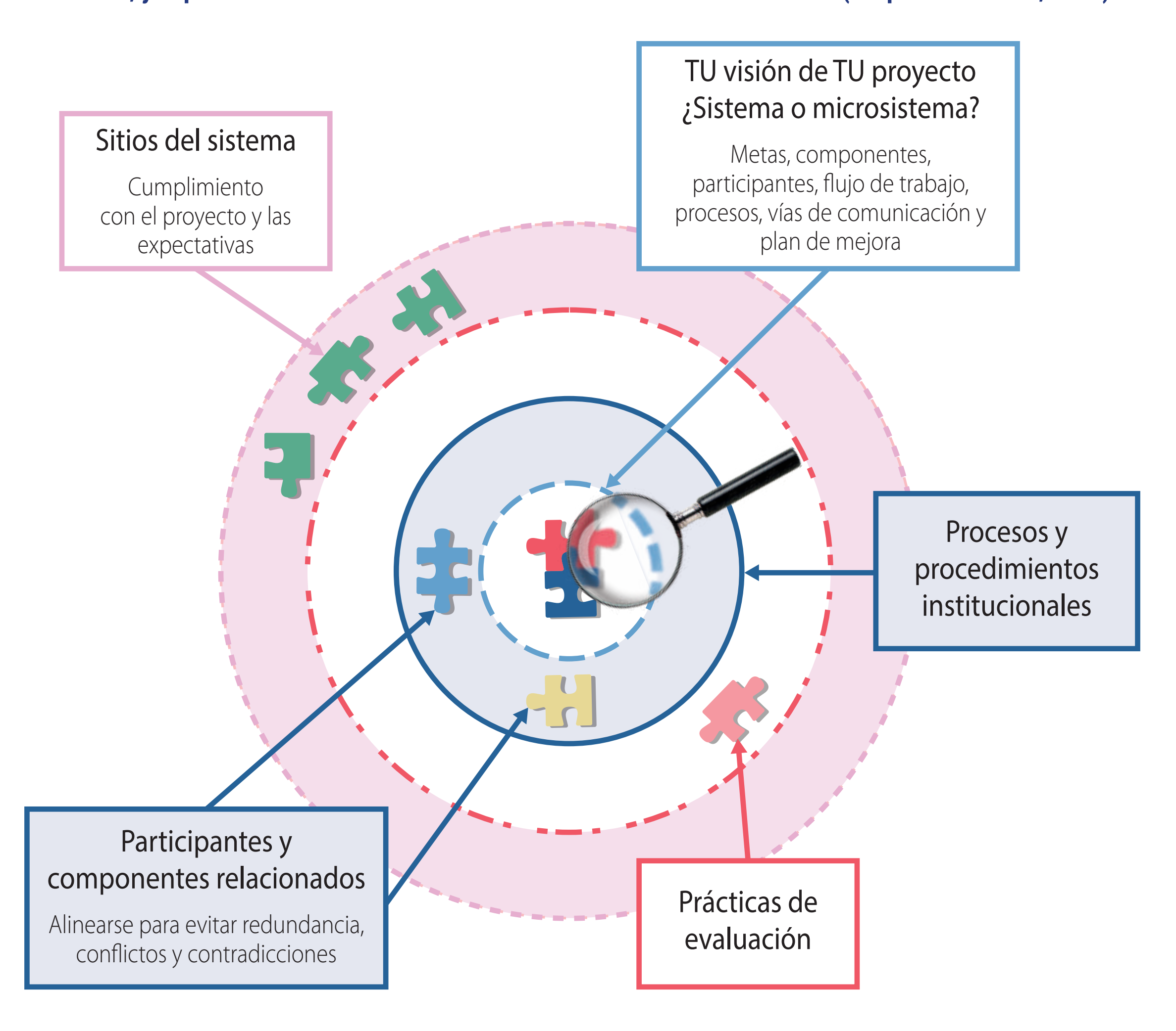
Tabla 2. Diseño de un programa educativo bajo la perspectiva de pensamiento de sistemas (adaptado de Bowe y Armstrong, 2017)

Diseño genérico de un sistema	Características de diseño de un programa educativo
Metas: definir especificaciones y capacidades del producto y del servicio. Asegurar que sean prioritarias.	**Metas educativas:** conocimientos, habilidades y competencias esperadas de los graduados del programa.
Flujo de trabajo: secuencia de etapas de desarrollo del producto y del servicio; resultados esperados de cada etapa de desarrollo.	**Currículo:** secuencia de unidades curriculares, cursos o rotaciones; resultados de aprendizaje esperados al final de cada unidad.

Diseño genérico de un sistema	Características de diseño de un programa educativo
Roles y responsabilidades de los participantes: contribuciones esperadas de los participantes en el desarrollo del servicio y del producto, de acuerdo con procesos y procedimientos definidos. Alinear los esfuerzos de las partes.	**Expectativas de estudiantes, profesores y directivos:** responsabilidades de aprendizaje, docencia y evaluación; resultados de aprendizaje detallados; criterios de desempeño de acuerdo con procesos y procedimientos.
Contexto estructurado: en el que se realiza el trabajo; se establecen procesos y procedimientos para facilitar conexiones, comunicación, coordinación y retroalimentación entre los participantes y elementos del sistema, atendiendo estándares definidos. Se intercambia información crítica para el desempeño del sistema como un todo. Se proporciona información oportuna, relevante y accionable para acciones correctivas, mejoría de calidad y capacidad de responder a cambios en ambientes interno y externo.	**Contexto del programa:** en el que ocurre la enseñanza y el aprendizaje; procesos y procedimientos planeados para facilitar la comunicación, retroalimentaciones y conexiones entre estudiantes, docentes y directivos, atendiendo estándares definidos. Se intercambia información crítica para el desempeño educativo del sistema, de forma oportuna y pertinente. Se realizan acciones correctivas resultado de la evaluación formativa, respondiendo a modificaciones en el contexto local y externo (universidad, sociedad).
Recursos y servicios del sistema: servicios de apoyo como tecnologías de información, manejo y análisis de datos, entrenamiento sobre el trabajo y planeación de carrera, proveedores de servicios.	**Recursos y servicios del sistema:** Ambientes de apoyo educativo, servicios y recursos; mentoría, consejería, formación docente, tecnologías de información, manejo y análisis de datos, investigación y apoyo de bibliotecas, becas, planeación de carrera, rotaciones en escenarios fuera del campus.

Al aplicar cada uno de los elementos de diseño de los sistemas y sus propósitos funcionales es que se pueden diseñar sistemas nuevos o evaluar los ya existentes. Por ejemplo, para el caso particular de evaluación del curso masivo abierto en línea descrito anteriormente, es necesario reflexionar: ¿cuáles son sus elementos y actores fundamentales?, ¿en dónde se encuentran los puntos de conexión de los elementos de ese sistema?, ¿qué fallas y áreas de oportunidad tiene ese sistema?, ¿cómo podría mejorarse? Para dar respuestas a estas interrogantes, se debe pensar en el modelo mental del sistema de evaluación que se desee desarrollar e implementar en su contexto particular. Es importante precisar que un modelo es una vista simplificada de una realidad compleja, implicando la creación de una abstracción que permite entender mejor un dominio (Eriksson y Penker, 2000). En la Figura 2 se muestran los diferentes elementos que conforman un sistema, en el que destacan las diferentes fronteras que hay entre ellos y cómo se pueden configurar como microsistemas, es decir, sistemas dentro de otros sistemas.

Figura 2. Ejemplo del modelo mental de un sistema con sus diferentes elementos, fronteras entre los mismos, y la posibilidad de tener microsistemas dentro de un macrosistema (adaptado de Bowe, 2019).



LAS CULTURAS Y SUBCULTURAS DE LOS SISTEMAS

Es importante mencionar que cada sistema posee un conjunto de culturas y subculturas dentro del mismo, que determinan lo que se puede hacer, las prioridades, las necesidades de cada uno de ellos. En la visión sistémica de la evaluación, es importante tener en mente a todos los participantes, directivos, docentes, estudiantes, recursos financieros y materiales, entre muchos otros. Mientras más se tengan en consideración estos aspectos, mayor posibilidad habrá de que el sistema funcione para lo que fue diseñado (Figura 3).

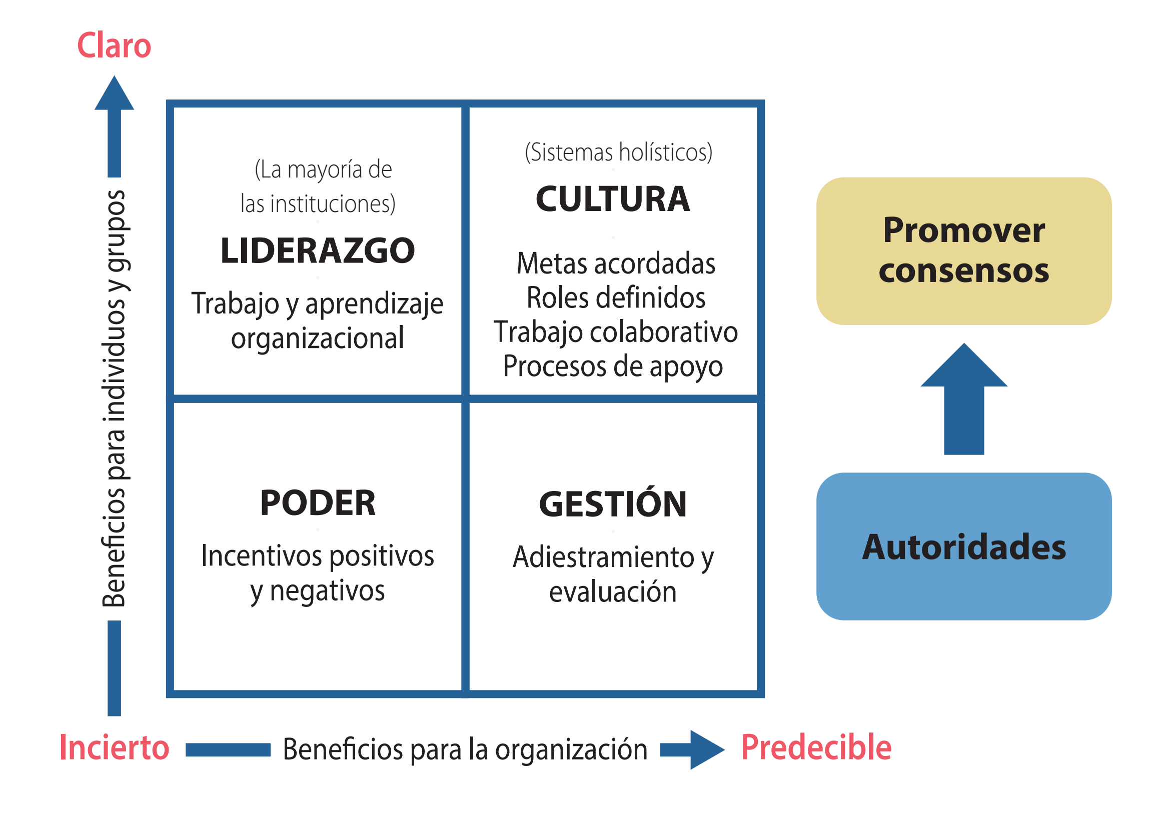
Figura 3. Subculturas organizacionales en la visión sistémica de la evaluación (adaptado de Bowe, 2019)



LOS BENEFICIOS DE LOS SISTEMAS

Como se puede ver en la Figura 4, si se grafica en el eje horizontal qué tan predecibles son los beneficios del sistema para la organización, y en el eje vertical qué tan claros son los beneficios para los individuos y el grupo, se forman cuatro cuadrantes que tienen que ver con liderazgo, la cultura de la organización, la gestión y el ejercicio del poder de los líderes asignados de la organización. El menos efectivo es la aplicación del poder sin tomar en cuenta lo demás, y el más claro y predecible es el que visualiza la cultura de la organización dentro de un sistema holístico, con los elementos que se comentaron anteriormente. Es muy importante promover consensos desde la fase de diseño de un sistema de evaluación, y construirlo en comunidad.

Figura 4. Predictibilidad de los beneficios de los sistemas para la organización versus individuos y grupos (adaptado de Stevenson, 1998)



LAS CLAVES DE LA EVALUACIÓN

Los roles funcionales claves de la evaluación en los sistemas son los siguientes:

- Proporcionar información para acciones correctivas, mejoría de calidad y planeación.
- Incentivar consideraciones incluyentes de los participantes, de factores que contribuyen a resultados de componentes y desempeño global del sistema, para avanzar a un modelo mental colectivo y aprendizaje organizacional.
- Apoyar la toma de decisiones basadas en evidencias.

El "hacer sentido" de los procesos influye en que avance el aprendizaje de toda la organización, aprovechando desde el nivel de datos hasta el desarrollo de la sabiduría individual y colectiva (Figura 5).

Figura 5. Elementos que constituyen el aprendizaje de las organizaciones (adaptado de Bowe, 2019)

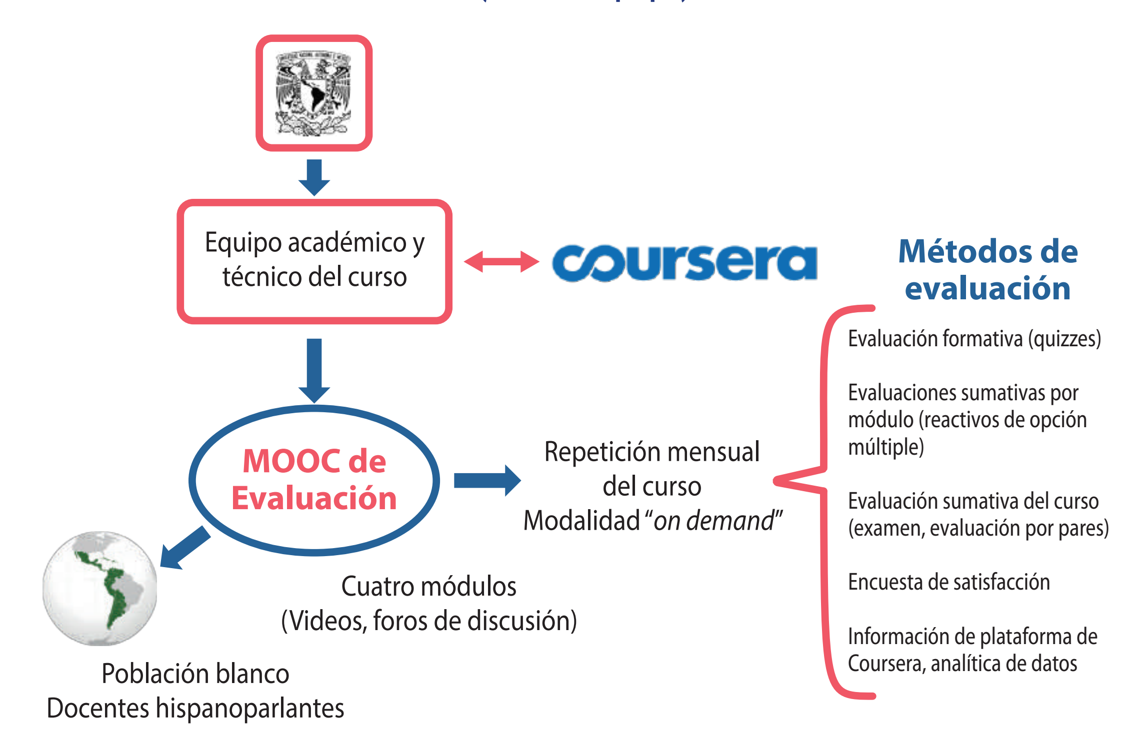
Es importante puntualizar que los datos son la base del aprendizaje organizacional, por lo que es vital desarrollar un buen nivel en términos de alfabetización de datos. De acuerdo con D'Ignazio y Bhargava (2015), describen la alfabetización de datos como la capacidad de:

- Leer datos, lo que significa comprender qué son los datos y qué aspecto del mundo representan.
- Trabajar con datos, incluida su creación, adquisición, limpieza y gestión.
- Analizar datos, lo que implica filtrar, clasificar, agregar, comparar y realizar otras operaciones analíticas con ellos.
- Discutir con datos, lo que significa usar datos para respaldar una narrativa más amplia que tiene la intención de comunicar algún mensaje o historia a una audiencia en particular.

EJEMPLO DE EVALUACIÓN SISTÉMICA DE UN MOOC

En el ejercicio de evaluar el curso abierto, masivo y en línea mencionado previamente, surge la pregunta: ¿cómo transformar la evaluación tradicional en una evaluación sistémica? En la Figura 6 se presenta un esquema tradicional de evaluación del curso.

Figura 6. Elementos y relaciones que constituyen un esquema tradicional de evaluación de un MOOC (elaboración propia)



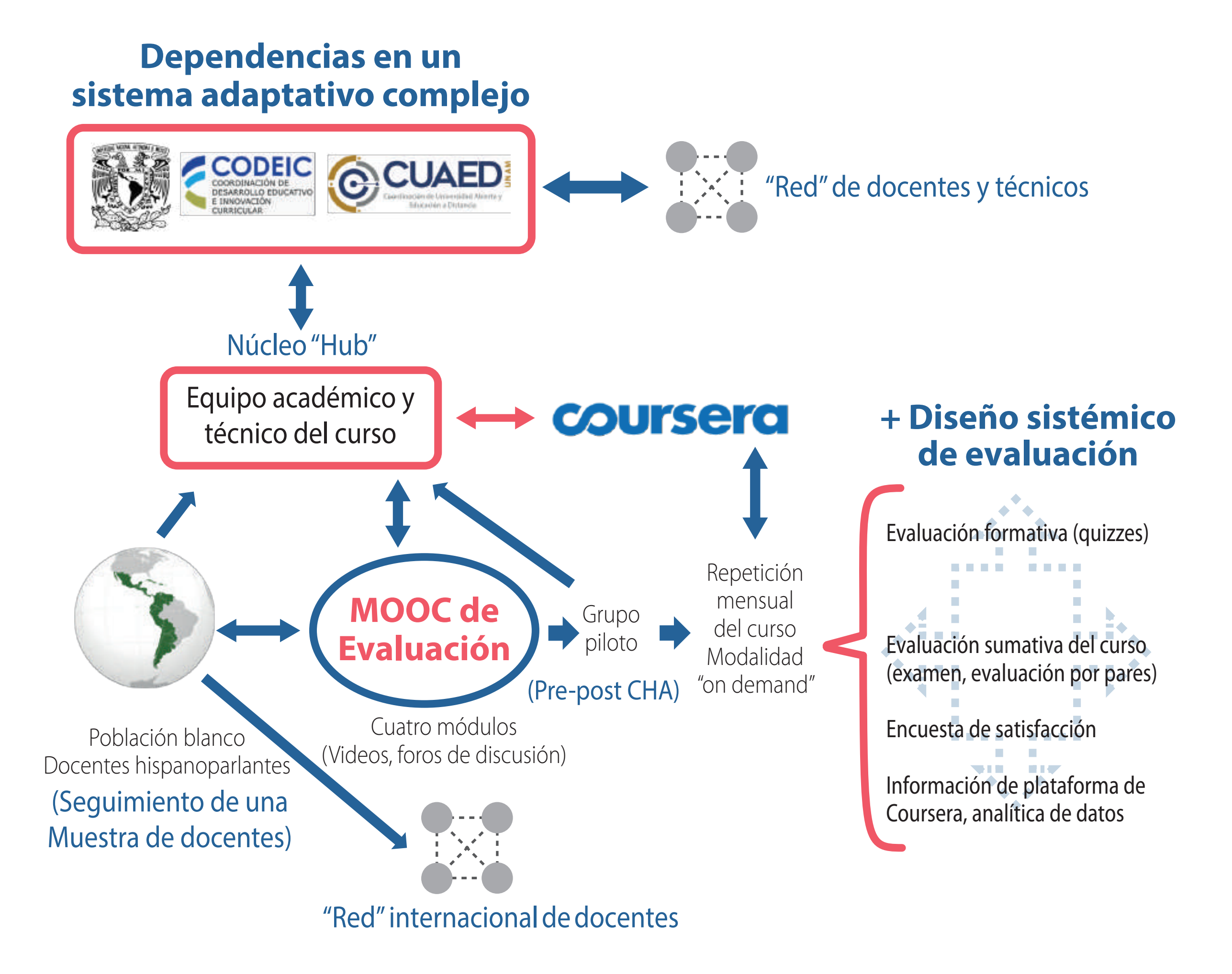
Para mejorar este esquema habría que hacer lo siguiente:

- Desarrollar un bosquejo del proyecto como un sistema, un mapa mental holístico que incluya los elementos críticos del mismo.
- Identificar los actores del proceso y las interacciones posibles entre ellos.

- Moverse del nivel microsistema a un nivel superior en que se visualice la relación entre sus componentes.
- Navegar entre lo micro y lo macro sin perder de vista el éxito del proyecto.

En la Figura 7 podemos ver un sistema más complejo y sofisticado para evaluar el MOOC, que incluye la interacción de diversos actores en lo que se llama un "sistema adaptativo complejo", con la presencia de núcleos de actividades y líneas de interacción multidireccionales, que incluye el uso de diversos instrumentos y estrategias evaluativas. Si se percibe con una perspectiva sistémica la evaluación desde un principio, el resultado será mejor y el proceso más satisfactorio y motivador para los participantes.

Figura 7. Elementos y relaciones que constituyen el sistema de evaluación del MOOC (CHA = conocimientos, habilidades, actitudes), (elaboración propia)



CONCLUSIONES

Se han descrito cinco reglas sencillas para entender y mejorar un sistema en educación (Armstrong et al., 2004):

1) Definir los productos esperados del sistema, lo que en educación universitaria es un reto complejo.
2) Definir un sendero o camino lo más claro y transparente posible, lo cual implica atención detallada a los aspectos curriculares y su alineamiento con los métodos de enseñanza y evaluación (Khanna et al., 2021).
3) Identificar las conexiones entre los diferentes elementos del sistema, para optimizar la comunicación entre las diferentes fases o unidades del sistema. Ello implica atención a la integración de las asignaturas, cursos, etapas del currículo, a través de mecanismos de comunicación y retroalimentación formales e informales.
4) Especificar las actividades del sistema, atender lo que ocurre de forma proactiva y en tiempo real, para mejorar la comunicación y desempeño de todo el sistema y sus microsistemas.
5) Asegurar la mejora continua como una actividad integrada por diseño en el sistema.

En resumen, podríamos enumerar las siguientes conclusiones:

- El enfoque sistémico de la evaluación permite tener una visión panorámica y al mismo tiempo profunda de todo el proceso.
- Contribuye a que no nos perdamos en los detalles y mantengamos la meta como "brújula".
- Proporciona información accionable a todos los elementos del sistema (docentes, estudiantes, tomadores de decisiones).
- No es sencillo, requiere actitud reflexiva, pensamiento crítico, capacidad de abstracción, liderazgo y trabajo colaborativo.
- Permite integrar los resultados de diferentes métodos e instrumentos en un todo coherente.

REFERENCIAS

Ackoff, R. L., and Emery, F. E. (1972) *On Purposeful Systems*. London:Tavistock Publications.

American Medical Association. (2019). Why you need to be a systems thinker in health care. American Medical Association. Consultado 26 de febrero 2022, en: https://www.ama-assn.org/education/accelerating-change-medical-education/why-you-need-be-systems-thinker-health-care

Aparicio, J., Rodríguez, D. Y., & Zabala-Iturriagagoitia, J. M. (2021). The systemic approach as an instrument to evaluate higher education systems: Opportunities and challenges. *Research Evaluation*, *30*(3), 336-348. https://doi.org/10.1093/reseval/rvab012

Armstrong, E. G., Mackey, M., & Spear, S. J. (2004). Medical education as a process management problem. *Academic Medicine: Journal of the Association of American Medical Colleges*, *79*(8), 721–728. https://doi.org/10.1097/00001888-200408000-00002

Bowe, C. M., & Armstrong, E. (2017). Assessment for Systems Learning: A Holistic Assessment Framework to Support Decision Making Across the Medical Education Continuum. *Academic Medicine: Journal of the Association of American Medical Colleges*, *92*(5), 585–592. https://doi.org/10.1097/ACM.0000000000001321

Bowe, C. M. (2019). "Systems Assessment Course". Harvard Medical International.

Colbert, C. Y., Ogden, P. E., Ownby, A. R., & Bowe, C. (2011). Systems-based practice in graduate medical education: systems thinking as the missing foundational construct. *Teaching and Learning in Medicine*, *23*(2), 179–185. https://doi.org/10.1080/10401334.2011.561758

Eriksson, H. E., & Penker, M. (2000). *Business modeling with UML*. New York, 1-12.

Dawidowicz, P. (2012). The person on the street's understanding of systems thinking. *Systems Research and Behavioral Science*, *29*(1), 2–13. https://doi.org/10.1002/sres.1094

D'Ignazio, C., & Bhargava, R. (2015). Approaches to Building Big Data Literacy. In Bloomberg Data for Good Exchange 2015. New York, NY, USA. https://www.media.mit.edu/publications/approaches-to-building-big-data-literacy/

Khanna, P., Roberts, C., & Lane, A. S. (2021). Designing health professional education curricula using systems thinking perspectives. *BMC Medical Education*, *21*(1), 20. https://doi.org/10.1186/s12909-020-02442-5

Ndaruhutse, S., Jones, C., Riggall, A. (2019). *Why systems thinking is important for the education sector.* Education Development Trust, Berkshire: United Kingdom. https://www.educationdevelopmenttrust.com/our-research-and-insights/research/systems-thinking-education-sector

Norman C. D. (2013). Teaching systems thinking and complexity theory in health sciences. *Journal of Evaluation in Clinical Practice*, *19*(6), 1087–1089. https://doi.org/10.1111/jep.12065

Plack, M. M., Goldman, E. F., Scott, A. R., Pintz, C., Herrmann, D., Kline, K., Thompson, T., & Brundage, S. B. (2018). Systems Thinking and Systems-Based Practice Across the Health Professions: An Inquiry into Definitions, Teaching Practices, and Assessment. *Teaching and Learning in Medicine*, *30*(3), 242–254. https://doi.org/10.1080/10401334.2017.1398654

Plack, M. M., Goldman, E. F., Scott, A. R., & Brundage, S. B. (2019). *Systems thinking in the healthcare professions: A guide for educators and clinicians*. Washington, DC: The George Washington University. https://hsrc.himmelfarb.gwu.edu/cgi/viewcontent.cgi?article=1000&context=educational_resources_teaching

Richmond, B. (1993). Systems thinking: Critical thinking skills for the 1990s and beyond. *System Dynamics Review*, *9*(2), 113–33. https://doi.org/10.1002/sdr.4260090203

Rosnay, J. (1977) *El Macroscopio*. Madrid: Editorial AC.

Senge, P. (2006). *The fifth discipline*. New York: Currency Doubleday.

Stave, K., & Hopper, M. (2007). What constitutes systems thinking? A proposed taxonomy. En: *Proceedings of the 26th International Conference of the System Dynamics Society*. Athens. http://static.clexchange.org/ftp/conference/CLE_2010/CO2010-06Session5MeasureST.pdf

Stevenson, H. H., Cruikshank, J. L., & Moldoveanu, M. C. (1998). Do lunch or be lunch: *The power of predictability in creating your future*. Harvard Business Press.

Sección II.

INSTRUMENTOS Y ESTRATEGIAS DE EVALUACIÓN (ENFOQUE CUANTITATIVO)

Capítulo 12

EXAMEN OBJETIVO

Adrián Martínez González, Manuel García Minjares, Careli Johana Herrera Penilla

"Technology will not replace great teachers
but technology in the hands of great teachers can be transformational."
George Couros

INTRODUCCIÓN

Hablar del proceso de enseñanza aprendizaje implica considerar elementos como los docentes, estudiantes, contextos educativos, contenidos, estrategias didácticas y evaluaciones, por citar algunos. Dentro de todos ellos, la evaluación es fundamental porque hace posible obtener evidencia del alcance del cumplimiento de los objetivos educativos durante y al término del proceso de enseñanza aprendizaje y contribuye en la mejora de la calidad del aprendizaje, por lo que es imperioso que se considere una actividad sistemática y continua capaz de recoger suficiente evidencia de validez de cada una de las partes que integran el proceso.

En el contexto de la educación presencial, uno de los instrumentos de evaluación de mayor uso es el examen objetivo o simplemente examen, el cual, correctamente elaborado, puede ser una herramienta de evaluación formativa o sumativa respecto a los contenidos o habilidades consideradas en el programa académico de estudios. Para su diseño, es necesario llevar a cabo una serie de pasos, así como tomar en cuenta ciertos requerimientos con la finalidad de que en realidad sea un instrumento objetivo que arroje suficiente evidencia de validez.

Este capítulo pretende brindar los elementos básicos para la conceptualización, elaboración y aplicación del examen objetivo conformado por reactivos de opción múltiple, en aras de mejorar la calidad del proceso de enseñanza aprendizaje en ambientes presenciales, a distancia o mixtos. El capítulo se organiza en dos secciones: la primera, detalla todo lo referente al concepto, características, ventajas, limitaciones, planeación, diseño y aplicación del examen objetivo en la modalidad presencial, así como de los diversos formatos que se pueden emplear en los reactivos de opción múltiple que lo integran; la segunda, describe, con base en la investigación reciente, lo referente al examen objetivo en el sistema de educación a distancia, con especial énfasis en su planeación, diseño y aplicación; finalmente, se incluyen

ejemplos, conclusiones y recomendaciones, así como una serie de recursos en línea y referencias que la interesada o interesado puede utilizar para construir adecuadamente su instrumento de evaluación con miras a fortalecer el proceso de enseñanza aprendizaje, incluso en un contexto de educación a distancia.

¿QUÉ ES?

El examen objetivo es un instrumento de evaluación que se presenta en forma escrita y está integrado por una variedad de reactivos; se le denomina objetivo debido a que solamente existe una respuesta correcta para cada uno de los reactivos que lo integran y el examinado debe elegirla entre un conjunto de opciones, lo cual impide que el juicio o las opciones del evaluador interfieran al momento de analizar los resultados (Burruezo Ordoñez et al., 2014).

Algunas de sus características son:

- Objetividad: Únicamente poseen una respuesta correcta, lo cual implica que al obtener los resultados no existe la intervención del juicio del evaluador.
- Validez: Es el grado en el que la evidencia y la teoría apoyan las interpretaciones de los resultados de una prueba para los usos propuestos del examen. Al realizar un análisis estadístico de los resultados, se puede detectar si la revisión cualitativa de la prueba corresponde con los resultados que arroja. Hay cinco fuentes de evidencia de validez: el contenido de la prueba, el proceso de respuesta, la estructura interna, su relación con otras variables y las consecuencias para la persona que es objeto de la evaluación (AERA, APA y NCME, 2018).
- Confiabilidad: Forma parte de la validez y se refiere a la precisión y permanencia con la que se obtienen los mismos resultados en diferentes poblaciones en condiciones semejantes esto quiere decir que al aplicar el mismo instrumento en diversas poblaciones los resultados estadísticos deben ser iguales o muy semejantes (AERA, APA y NCME, 2018).

Ventajas:

- Tiene criterios de calificación unívocos y precisos, lo cual permite que los resultados se obtengan de forma rápida y confiable (Morales Vallejo, 2006).
- Con él, es posible medir el nivel de conocimiento, la comprensión o la capacidad para aplicar algún conocimiento específico que posee un individuo para resolver determinado problema.
- Los reactivos que los integran son una muestra representativa y equilibrada de los conocimientos que se desea evaluar, lo cual permite valorar uno o más temas o subtemas en un solo examen (Downing y Haladyna, 2006).
- Permite comparar el desempeño de individuos y grupos (Downing y Haladyna, 2006).
- Es un instrumento flexible que cuenta con una aparente facilidad de elaboración, aplicación y análisis de los resultados.

Limitaciones:

- Es posible contestar al azar y acertar a la respuesta correcta, lo cual no quiere decir que los estudiantes, efectivamente posean el conocimiento, sin embargo, en un examen objetivo elaborado apegándose a una metodología ordenada y sistemática esta limitación prácticamente no existe (Abedalqader et al., 2015).
- Se puede caer en evaluar principalmente la memoria o la recuperación de información al elaborar los reactivos, en ese caso estaríamos evaluando únicamente esta habilidad en los estudiantes.
- Puede existir un entrenamiento previo en los estudiantes para responder por inferencia este tipo de exámenes, de acuerdo con la estructura de las opciones de los reactivos de opción múltiple (Phelps, 2017).
- Por medio de ello, solo se puede conocer el resultado final de los aprendizajes esperados, pero no el proceso por medio del cual se llegó a ese resultado.
- Requieren de una retroalimentación después de los resultados para conocer el porqué de los aciertos y errores, esto con la finalidad de reforzar los aprendizajes esperados.

Reactivos de opción múltiple (ROM)

Un reactivo es una unidad de medida que consiste en una pregunta o instrucción que requiere una respuesta por parte del examinado y a partir de ella se puede inferir su ejecución o desempeño en el objeto de la evaluación. (Haladyna et al., 2002).

Los elementos que conforman a un reactivo de opción múltiple generalmente son los siguientes:

- Tallo o base: También conocido como enunciado, es el planteamiento que da lugar a una respuesta por parte del examinado. Dentro de la base o tallo se puede incluir material gráfico, como son ilustraciones, tablas, fotografías y esquemas, también puede ser un párrafo de alguna obra literaria o algún artículo de investigación.
- Respuesta correcta: Consiste en la opción de respuesta que corresponde con lo que se solicita en el tallo o base del reactivo.
- Distractor(es): Son aquellas opciones de respuesta que tienen una concordancia gramatical y semántica con el tallo o base del reactivo, pero que son totalmente incorrectas de acuerdo con lo que se solicita en él.

Tipos de reactivos de opción múltiple:

- Cuestionamiento directo: Son aquellos en cuyo tallo o base se plantea una afirmación o pregunta y se debe elegir la opción que responda correctamente a ese planteamiento.
- Completamiento: Son aquellos en cuyo tallo o base se presentan entre una y tres líneas vacías, a fin de que estas se completen de manera correcta al elegir la opción que contenga los elementos que le faltan al tallo o base.
- Apareamiento o relación de columnas: Son aquellos cuya base se conforma por dos

componentes: una instrucción y dos conjuntos de elementos. La primera indica cómo se debe realizar la relación entre dichos conjuntos (letras con números, generalmente), para posteriormente, elegir la opción que presenta la respuesta correcta.

- Jerarquización u ordenamiento: Son aquellos cuyo tallo o base se conforma por dos componentes, una instrucción que solicita ordenar de manera jerárquica o cronológica, y un conjunto de elementos que se enlistan de manera desordenada, para posteriormente elegir la opción que presenta el orden correcto.
- Resolución de problemas: Son aquellos en cuyo tallo o base se plantea un problema que requiere de un proceso para su resolución, para posteriormente, elegir la opción correcta que presenta únicamente el resultado final de ese proceso.
- Selección o complementación múltiple: Son aquellos cuyo tallo o base se compone de dos partes, una afirmación o pregunta y una lista de elementos, de los cuales, solo algunos son correctos como respuesta, por tanto, se debe elegir la opción que contiene los que son correctos.

¿CÓMO LO DISEÑO?

La planeación es el momento en el que se puntualizan una serie de acciones que, llevadas a cabo de manera seriada y ordenada, permiten alcanzar cierto objetivo, en este caso, es la construcción de un examen objetivo con reactivos de opción múltiple, para lo cual es importante considerar dos aspectos, las características del examen en sí mismo y las características de los reactivos de opción múltiple que lo integrarán; es lo que a continuación se describe (Downing y Haladyna, 2006).

Planeación

1) Elabore un “Perfil de referencia”.
   Es un documento en el que se especifica lo siguiente: título y tipo de examen, propósito del examen y temas, contenidos o conocimientos a evaluar. Este documento se puede elaborar a partir de la respuesta a la pregunta: ¿de mi asignatura, qué temas, contenidos o conocimiento quiero evaluar y con qué objetivo? (Consulte el “Ejemplo 1. Perfil de referencia”)
2) A partir de los contenidos o temas seleccionados, elabore una “Tabla de especificaciones”.
   Es un documento en el que se señalan los resultados de aprendizaje, que estarán asociados a cada tema o subtema que se incluirá en el examen. Los resultados de aprendizaje son afirmaciones que expresan lo que el alumno debe demostrar que sabe; deben ser muy específicos, solo deben evaluar un tema o subtema a la vez y puede establecerse más de un resultado de aprendizaje por tema o subtema. Este documento se puede desarrollar dando respuesta a la pregunta: ¿qué espero que los estudiantes demuestren que saben de determinado tema o subtema?
3) Otorgue el peso específico o ponderación para cada tema en la “Tabla de especificaciones”.
   El peso específico o ponderación para cada tema contenido en la “Tabla de especificaciones”,

se otorga de acuerdo con el tiempo destinado en clases para revisar los temas, o bien, de acuerdo con la importancia que el profesor considera que cada tema tiene en la formación de los estudiantes. El total de la ponderación en la "Tabla de especificaciones" debe sumar 100% que es a lo que equivaldrá el examen. (Consulte el "Ejemplo 2. Tabla de especificaciones")

4) Defina el número de reactivos requeridos para evaluar cada tema o subtema.
   Para lograr el balance adecuado de contenidos evaluados en el examen, se debe agregar a la "Tabla de especificaciones" el número de reactivos necesarios por tema a partir del peso específico o la ponderación previamente establecida.
5) Elabore los reactivos para cada resultado de aprendizaje.
   Ya que se cuenta con el número de reactivos requeridos por tema o subtema y con los resultados de aprendizaje asociados a estos, se deben elaborar los reactivos de opción múltiple que se requieren para integrar el examen. (Revise la sección "Elaboración de buenos reactivos de opción múltiple")
6) Si es posible, integre y aplique un examen piloto.
   Un examen piloto está integrado por reactivos nuevos y se aplica para evaluar su funcionamiento para posteriormente, realizar modificaciones en ellos a fin de ser utilizados en un examen definitivo.
7) Calibre los reactivos que se aplicaron en el examen piloto.
   La calibración es un proceso cuantitativo donde se calculan diversos parámetros que permiten evaluar tanto el funcionamiento de los reactivos como del instrumento. Existen algunas alternativas de *software* libre que permite realizar la calibración de un examen como lo es el Sistema de Análisis Psicométrico de Reactivos (SISAPRE) desarrollado por la Coordinación de Universidad Abierta Innovación Educativa y Educación a Distancia de la UNAM.
8) Integre una versión del examen definitivo.
   A partir de la "Tabla de especificaciones" y considerando aquellos reactivos que hayan logrado estándares psicométricos adecuados, seleccione los reactivos que integrarán el examen final o definitivo para ser aplicado.
9) Integre y resguarde la cadena de respuestas del examen definitivo.
   Es importante que una vez que el examen definitivo quede integrado, se resguarde la cadena de respuestas correctas, esto agilizará el proceso de calibración final y calificación.
10) Comience a integrar un "Banco de reactivos".
   Un "Banco de reactivos" consiste en concentrar los reactivos nuevos y previamente utilizados en un solo espacio, ya sea virtual o físico, con la finalidad de almacenarlos de manera ordenada y clasificarlos por tema o subtema para poder ensamblar diversas versiones de exámenes, pues para obtener mejores resultados en las evaluaciones por medio del examen objetivo, es recomendable no utilizar los mismos reactivos más de dos veces seguidas; es conveniente variar los reactivos que lo integran, se puede ensamblar un examen utilizando reactivos nuevos y previamente utilizados.

Elaboración de buenos reactivos de opción múltiple

1) Identifique y comprenda lo que solicita el resultado de aprendizaje.
   Esto puede realizarse muy bien respondiendo a la pregunta: ¿qué es lo que se espera que demuestre el examinado al responder el reactivo?
2) Defina el tipo de reactivo que es más apropiado para ese resultado de aprendizaje.
   Entro los tipos de reactivos de opción múltiple que se han explicado previamente, es necesario, determinar ¿cuál es el más adecuado para elaborar el reactivo que estará asociado a ese resultado de aprendizaje?
3) Redacte el tallo o base del reactivo.
   Defina si esta será una afirmación, una pregunta, una instrucción, etcétera y si contendrá únicamente texto o también se requieren imágenes, esquemas, gráficas, entre otros elementos.
4) Redacte la respuesta correcta.
   Esta debe ser clara y sin lugar a duda la única opción correcta; no debe ser parcialmente correcta ni sugerir pistas para su elección.
5) Redacte el resto de las opciones o distractores.
   Se recomienda que sean tres distractores, que estos correspondan al tema o subtema y campo semántico, tanto del tallo o base del reactivo, como de la respuesta correcta y que tengan una concordancia de género y número con el tallo o base.

Ahora bien, con la finalidad de garantizar que los reactivos elaborados cumplen con los criterios de elaboración, se sugiere utilizar la lista de cotejo desarrollada por Rivera Jiménez et al. (2016) la cual contiene criterios específicos que ayudarán a verificar la validez de los reactivos que se elaboren.

Tabla 1. Lista de cotejo para la verificación de validez de un reactivo

Factores a considerar	Criterios de validez	Cumple	No cumple
Comprensión del reactivo	¿La cantidad de texto en la base es adecuada para su comprensión?		
	¿El tallo o base se encuentra redactado con claridad?		
	¿El reactivo cuenta con una gramática, puntuación y ortografía correcta?		
	¿La base del reactivo plantea la idea central?		
Contenido del reactivo	¿El reactivo mide un solo resultado de aprendizaje?		
	¿El reactivo presenta un solo contenido temático?		
	¿La semántica utilizada está de acuerdo con el contenido del programa académico?		

Factores a considerar	Criterios de validez	Cumple	No cumple
Precisión del reactivo	¿El reactivo cuenta únicamente con una respuesta correcta?		
	¿Las opciones de respuesta son independientes entre sí?		
	¿El contenido evaluado está en relación con la especificación del reactivo?		
Redacción de opciones de respuesta	¿Las opciones de respuesta son similares en cuanto a estructura gramatical, contenido y extensión?		
	¿Las opciones de respuesta evitan dar pistas sobre la respuesta correcta?		
	¿Los distractores son plausible, es decir, no se descartan por inferencia lógica o sentido común?		
	¿El reactivo cuenta con tres o cuatro opciones de respuesta?		

¿CÓMO LO APLICO?

- Es deseable que la aplicación sea en un espacio ventilado, limpio y correctamente iluminado, así como contar con el tiempo suficiente para responder a cada uno de los reactivos.
- Al inicio de la aplicación se deben dar las instrucciones claras y precisas para poder responder el examen. Es importante establecer el tiempo con que se cuenta para responder, resaltar la importancia de colocar su nombre completo en él, así como, si se debe contestar con alguna herramienta en específico (lápiz, pluma, etcétera) y si se puede hacer uso de algún apoyo para responderlo (calculadoras, tablas periódicas, traductoras, *tablets*, cuadernos de apuntes, etcétera.)
- Se recomienda entregar el examen al terminar de dar las instrucciones, pues generalmente, si se entrega antes, los examinados comienzan a revisarlo y no ponen atención a las indicaciones.
- Se debe mencionar en qué modalidad será el examen, es decir, si se resolverá en equipos o de manera individual, así como el objetivo de su aplicación.
- Es importante reiterarles a los examinados que deben leer con cuidado cada uno de los reactivos antes de responderlos.
- Durante la aplicación pueden surgir dudas o comentarios respecto a la redacción, orden, o contenidos de los reactivos, por ello, es recomendable pedirles a los examinados que, si surge cualquier tipo de duda o comentario durante la aplicación, levanten la mano para que de forma individual y sin alterar el orden del grupo, se hagan las aclaraciones pertinentes.
- En caso de que la duda o comentario afecte de forma general a las respuestas del examen, es decir, que exista algún error en alguno de los reactivos, es fundamental que se realice

una aclaración de forma grupal, o bien, solicitar que no respondan ese reactivo debido a que presenta un error.

- Mantener la disciplina en el grupo antes y durante la aplicación es benéfico en dos sentidos, por un lado, permite que los examinados comprendan y ejecuten adecuadamente las instrucciones que se les dieron para responder el examen, y por otro, facilita que la lectura y la concentración en los reactivos que conforman el instrumento sean adecuados para obtener mejores resultados.
- Es muy importante supervisar la aplicación, ya que es posible que los estudiantes tomen fotografías con algún dispositivo electrónico o bien, que sustraigan información del examen para filtrarla posteriormente.
- Se recomienda que conforme los examinados, vayan terminado de responder el examen, lo entreguen, y después de revisar que esté completo (datos de identificación del examinado y respuestas), se retiren de ese espacio para no generar ningún tipo de distracción a los examinados que aún se encuentran realizándolo.
- Al final de la aplicación, es deseable concentrar los exámenes en un solo lugar una vez concluida la aplicación y resguardarlos para su revisión y análisis.
- Es importante asegurarse de que ningún examinado extrajo el examen del lugar, o bien, tomó algunas fotografías de este, ya que es un instrumento de carácter confidencial.

EL EXAMEN OBJETIVO EN LA EVALUACIÓN A DISTANCIA

El uso del examen objetivo dentro del contexto de la educación a distancia como un instrumento de evaluación eficiente y confiable, ha estado presente aun antes de la pandemia por COVID-19, de esto dan cuenta varios artículos y libros que hablan del uso de este instrumento para la evaluación en contextos educativos diferentes al de la modalidad presencial (Centeno Brambila y Lira Obando, 2015; Dlab, M.H. et al., 2015; Universidad Autónoma de Chihuahua, 2012). Ante la necesidad de brindar educación únicamente a distancia debido a la pandemia, muchas instituciones de todos los niveles comenzaron a generar manuales y recomendaciones para poder aplicar este instrumento a distancia sin perder objetividad, validez y confiabilidad en sus resultados (Universidad de la República de Uruguay, 2020; Universidades Públicas de Castilla y León, 2020; UNESCO, 2020; UNAM-CUAIEED, 2020).

Además de seguir las diversas recomendaciones que varias instituciones educativas han considerado efectivas para aplicar este instrumento a distancia y en un ambiente no controlado obteniendo resultados óptimos, cabe resaltar el hecho de que es posible utilizar este instrumento de evaluación y obtener resultados válidos si se sigue una metodología adecuada y sobre todo si se elaboran de manera cuidadosa los reactivos de opción múltiple que lo conforman. También es importante mencionar que este instrumento no debe ser el único que se utilice para realizar evaluación en un contexto tan diverso como lo es la educación a distancia, es muy recomendable que se utilicen varios instrumentos, o bien, que este instrumento se utilice mucho más para realizar evaluaciones de carácter formativo, realizando sesiones de

retroalimentación con los estudiantes en las cuales se puedan comentar las fortalezas y áreas de oportunidad que se identifican fácilmente a partir de los resultados que arroja.

Si bien es cierto que, como puede suceder con cualquier instrumento de evaluación, en el caso del examen objetivo, aun no contamos con medios suficientes para garantizar que los estudiantes no incurran en "trampas", también es cierto, que ya contamos con recomendaciones y diversas herramientas tecnológicas que nos pueden ayudar a aminorar el impacto de esas acciones por parte de los estudiantes, en los resultados obtenidos (González, Marco y Medina, 2020).

¿CÓMO DISEÑO EL EXAMEN OBJETIVO A DISTANCIA?

- El diseño del examen objetivo en el contexto de la educación a distancia depende en gran medida del tipo de *software*, plataforma o aplicación que el docente desee utilizar, por ello, es importante conocer y utilizar varias de estas con la finalidad de encontrar la que más se adapte a las necesidades del contenido evaluado y los recursos con los cuales nuestros estudiantes cuentan para poder responder el examen.
- Para el diseño de un examen objetivo a distancia, es muy importante considerar el tipo de aplicación que se realizará, es decir, puede ser de manera sincrónica o asincrónica. No es lo mismo realizar un examen en una sesión por *Zoom* que pedirles a los estudiantes que respondan el examen en una plataforma que estará abierta permanentemente y a la cual ellos podrán acceder en cualquier momento para responder el examen.
- Aunque sea en un contexto de educación a distancia, también se debe elaborar un Perfil de referencia y una Tabla de especificaciones, dando clic sobre los enlaces encontrará dos videos breves en donde se describen más a detalle las características que deben tener ambos documentos.
- Para poder ensamblar el examen objetivo de una manera mucho más rápida, es recomendable que el *software*, plataforma o aplicación que se utilice cuente con un espacio lo suficientemente amplio para poder almacenar todos los reactivos que se vayan elaborando y utilizando, de esta manera es posible construir un banco de reactivos en el mismo espacio en donde se generen los exámenes y esto permite emplear menos tiempo al momento de ensamblar diferentes versiones de exámenes del mismo tema o subtema. Dando clic en el enlace DAEX se encuentra la plataforma "Diseño y Aplicación de Exámenes", desarrollada por la CUAIEED para apoyar a los docentes en la elaboración de exámenes objetivos de acuerdo con la metodología descrita anteriormente, lo cual garantiza mejores resultados en las evaluaciones por medio del examen objetivo.

¿CÓMO LO INSTRUMENTO?

La instrumentación de un examen objetivo en línea, se puede realizar de dos formas: sincrónica o asincrónicamente, pero lo que en ambos casos es indispensable, es que todos los involucrados en el proceso de aplicación (estudiantes, docentes y/o personal de apoyo técnico), conozcan a cabalidad, tres elementos: el primero son los requerimientos técnicos que son

indispensables para la resolución del examen, el segundo, son indicaciones claras y precisas para llevar a cabo el proceso de aplicación y el tercero, las reglas para no incurrir en alguna sanción que amerite la suspensión del examen durante su aplicación o bien, el no considerar su resultado, por todo esto, es recomendable realizar un documento en donde se especifiquen puntualmente estos tres elementos y socializarlo a fin de que un poco antes del examen se esclarezcan dudas respecto a estos y los estudiantes tengan la oportunidad de externar sus dudas o inquietudes, por ejemplo, que no cuentan con alguno de los requerimientos técnicos, este documento es muy importante pues tiene la finalidad de prevenir y solucionar cualquier inconveniente que se pueda presentar y de esta manera garantizar una aplicación exitosa.

Recomendaciones para aplicación sincrónica

- La aplicación sincrónica se refiere al uso de alguna herramienta tecnológica que permita al docente supervisar la aplicación del examen en tiempo real, algunas de ellas son *Zoom*, *Cisco Webex Meetings*, *Google Meet*, *Microsoft Teams*, entre otras.
- Es recomendable que el grupo de estudiantes al cual será aplicado el examen sea reducido, esto facilitará al docente tener un mejor control de la aplicación.
- Si no es posible contar con grupos reducidos para la aplicación, es recomendable, dividir al grupo y realizar dos o tres versiones del examen, con diferentes reactivos, pero del mismo contenido temático, de esta manera se podrá aplicar el examen a los grupos, minimizando la posibilidad de que se pasen las respuestas.
- Antes de iniciar el examen, se les puede solicitar a los estudiantes que muestren su espacio de trabajo y verificar que esté libre de cualquier cosa o persona que pueda distraerlos o de la cual puedan hacer uso para obtener las respuestas del examen.
- Durante el examen es recomendable que los estudiantes tengan encendidos cámara y micrófono, a fin de "simular" que se encuentran en el salón de clase y de que el docente pueda monitorear la aplicación del examen.
- Es recomendable que una vez que el estudiante termine de resolver el examen, dé aviso al profesor y no abandone la sesión, lo ideal es que permanezca conectado frente a cámara esperando al resto del grupo, esto con la finalidad de evitar que al abandonar la sesión pueda establecer comunicación con alguno de los estudiantes que aún se encuentran realizando el examen.
- Después de la aplicación es importante notificar a los estudiantes en cómo, dónde y cuándo podrán conocer su resultado.

Recomendaciones para aplicación asincrónica

- La aplicación asincrónica se refiere al uso de alguna herramienta electrónica que permita almacenar el examen a fin de que los estudiantes puedan tener acceso a este, responderlo y almacenar los resultados, todo esto sin supervisión del docente en tiempo real.
- Antes de la aplicación es recomendable cerciorarse de que todos los estudiantes saben cómo, dónde, cuándo y con cuánto tiempo cuentan para acceder, resolver y enviar el examen terminado.

- Durante la aplicación es importante estar al pendiente de los canales de comunicación que se han establecido con los estudiantes, como puede ser, correo electrónico, plataforma, *Whatsapp*, entre otros, esto con la finalidad de que, si surgen dudas durante el examen, puedan ser resueltas sin entorpecer el proceso de respuesta de los estudiantes.
- Al final de la aplicación es recomendable cerciorarse de que todos hayan resuelto el examen en tiempo y forma y si se detecta algo fuera de lo común, comunicarse lo más pronto posible con él o los estudiantes; no perder la comunicación durante todo el proceso de aplicación es muy importante para garantizar su éxito.

Actividad interactiva

Ahora realiza un breve quiz respecto al "Examen objetivo", por favor, ingresa a la siguiente liga https://quizizz.com/join e ingresa el código número 33980462 en el espacio que dice "Ingrese un código para unirse" y dé clic en el botón "ÚNETE". Realiza el quiz para verificar la obtención de los conocimientos adquiridos durante la lectura de este capítulo.

¿CÓMO ANALIZO SUS RESULTADOS?

- Los resultados obtenidos en este tipo de instrumento nos permiten medir el grado de conocimiento, comprensión o aplicación que han alcanzado los alumnos con respecto al contenido de la asignatura, módulo o curso en uno o varios temas.
- Al ser un instrumento de carácter objetivo y cuantitativo, es necesario obtener los resultados a partir de un conjunto de operaciones matemáticas que nos arrojen, no solo el grado individual de desempeño de los estudiantes, sino también del grupo en general y del comportamiento de cada uno de los reactivos que se usaron en la prueba; ello permite probar los reactivos con la finalidad de incrementar su precisión, validez y confiabilidad, es decir, que midan lo que se pretende medir con cada uno de ellos.
- El análisis de los resultados en el caso de aplicaciones masivas de exámenes objetivos se realiza por medio del uso de programas de *software*, los cuales permiten que, al introducir la cadena de respuestas correctas y leer las cadenas de cada uno de los examinados, este realice cálculos de carácter estadístico que arrojan datos, tales como: la dificultad de cada uno de los reactivos, el comportamiento de cada uno de los distractores en cada reactivo, la discriminación en las respuestas de los examinados, entre otros.
- Si se desean obtener los parámetros psicométricos de los exámenes, dando clic sobre el enlace SISAPRE se encuentra la plataforma "SISAPRE", desarrollada por la CUAIEED para apoyar a los docentes en la obtención de resultados de exámenes objetivos de opción múltiple.
- Es importante considerar que si el alumno sabe que solamente será evaluado con un examen objetivo que cubrirá ciertos temas, únicamente se esmerará en memorizar el contenido a evaluar, lo cual entorpece el proceso de aprendizaje. Por ello, es importante que el examen este balanceado a fin de evaluar conocimiento, comprensión y aplicación

y además de lo anterior, que este tipo de instrumento se complemente con algunos de carácter cualitativo, así se podrá tener mayor certeza de que el estudiante no solo memorizó los contenidos a fin de obtener una buena calificación en el examen.

EJEMPLOS

Ejemplo 1. Perfil de referencia

Primer examen parcial de la asignatura de Matemáticas I
Tipo de examen: Evaluación parcial.
Propósito del examen: Conocer el grado de conocimientos y comprensión de los estudiantes de los temas revisados durante los dos primeros meses del curso.
Institución: Universidad Nacional Autónoma de Nuevo León (UNANL)

Temas a evaluar:
1. Números reales.
2. Números completos.

Subtemas a evaluar:
- 1.1 Suma y resta con números reales.
- 1.2 Multiplicación y división con números reales.
- 2.1 Suma y resta con números complejos.
- 2.2 Multiplicación con números complejos.

Resultados de aprendizaje:
- 1.1.1 Calcula operaciones (suma y/o resta) con números reales en forma algorítmica.
- 1.1.2 Soluciona problemas con el desarrollo de operaciones (suma y/o resta) con números reales.
- 1.2.1. Calcula operaciones (multiplicación y/o división) con números reales en forma algorítmica.
- 1.2.2 Soluciona problemas con el desarrollo de operaciones (multiplicación y/o división) con números reales.
- 2.1.1 Calcula operaciones (suma y/o resta) con números complejos en forma algorítmica.
- 2.2.1. Calcula operaciones (multiplicación) con números complejos en forma algorítmica.

Fuente: Elaboración propia.

Ejemplo 2. Tabla de especificaciones

Asignatura: Matemáticas I **Examen:** Primer parcial **Elaborador:** Mauricio Cuevas		**Fecha de elaboración:** 03-04-2016 **Institución:** UNANL **Grupo:** 001		
Clave de temas y subtemas	Temas y subtemas	Resultados de aprendizaje	Ponderación	Número de reactivos
1.0.0.0	Números reales		50%	5
1.1.0.0	Suma y resta con números reales	1.1.1 Calcula operaciones (suma y/o resta) con números reales en forma algorítmica.		
		1.1.2 Soluciona problemas con el desarrollo de operaciones (suma y/o resta) con números reales.		
1.2.0.0	Multiplicación y división con números reales	1.2.1. Calcula operaciones (multiplicación y/o división) con números reales en forma algorítmica.		
		1.2.2 Soluciona problemas con el desarrollo de operaciones (multiplicación y/o división) con números reales.		
2.0.0.0	Números complejos		50%	5
2.1.0.0	Suma y resta con números complejos	2.1.1 Calcula operaciones (suma y/o resta) con números complejos en forma algorítmica.		
2.2.0.0	Multiplicación con números complejos	2.2.1 Calcula operaciones (multiplicación) con números complejos en forma algorítmica.		

Fuente: Elaboración propia.

Ejemplo 3. Tipos de reactivos

Cuestionamiento directo
Estructura que protege y da soporte a las células vegetales.

a) Complejo de Golgi.
b) Pared celular.
c) Retículo endoplásmico.
d) Membrana plasmática.

Completamiento
La novela pertenece al género __________, está escrita en _________ y suele tener una estructura _________.

a) narrativo – prosa – compleja
b) épico – verso – intensa
c) dramático – verso – constante
d) lírico – prosa – simple

Jerarquización u ordenamiento
Ordena en forma descendente las categorías taxonómicas.

I. Orden.
II. Género.
III. Familia.
IV. Especie.

a) II, IV, III y I
b) I, III, II y IV
c) III, I, II y IV
d) IV, II, III y I

Resolución de problemas
Un niño está parado sobre un carrito, ambos en reposo, el primero pesa 30 kg y cuando salta hacia adelante a 2 m/s, el carrito es lanzado hacia atrás a 12 m/s. Si se desprecia la fricción, ¿cuál es la masa del carrito?

a) 6.0 kg
b) 5.0 kg
c) 2.5 kg
d) 0.8 kg

Selección o complementación múltiple
Regiones densamente pobladas en el mundo.

I. Este de Europa.
II. Oeste de Oceanía.
III. Este de África.
IV. Norte de Europa.
V. Sur de Asia.

a) III y IV
b) I y II
c) III y V
d) I y V

Fuente: CUAIEED, 2020.

CONCLUSIONES Y RECOMENDACIONES

- Un examen objetivo correctamente elaborado puede ser utilizado como un instrumento de selección o diagnóstico de tipo formativo o sumativo en función de una asignatura, módulo o programa de estudios.
- Sus características principales son la objetividad, la validez y la confiabilidad.
- Se requiere elaborar un perfil de referencia y una tabla de especificaciones para comenzar a planearlo y elaborarlo correctamente, esto con la finalidad de determinar los aprendizajes esperados que se desean evaluar a través de este instrumento.
- Es muy importante mantener comunicación durante todo el proceso de aplicación en el contexto de la educación a distancia y supervisar a los estudiantes en el contexto de la educación presencial, pues cualquier descuido puede afectar directamente los resultados.
- Es muy importante evaluar de manera periódica y cualitativa y cuantitativa los reactivos que integran el examen, pues esto permite mejorar el instrumento progresivamente y de una manera más integral.

Tabla 3. Recursos en línea

Recurso	Descripción
Diseño y Aplicación de Exámenes (DAEX) www.daex.unam.mx	Es una plataforma diseñada para gestionar la elaboración y aplicación de exámenes de opción múltiple (formativos y sumativos), fácil de usar, novedosa y segura que pretende apoyar la labor docente y de investigación en la UNAM.
Sistema de Análisis Psicométrico de Reactivos (SISAPRE) www.sisapre.cuaieed.unam.mx	Es una plataforma destinada a apoyar a docentes e investigadores en la evaluación psicométrica de exámenes objetivos y responde a la necesidad de los responsables de evaluación de contar con un *software* especializado gratuito.
Crear Test www.creartest.com	Es un sitio web gratuito que permite crear exámenes y test on line y se pueden compartir en Facebok, Moodle, WhatsApp, Twitter y otras redes sociales.
Examprofessor www.examprofessor.com	Es un sitio web que permite crear y administrar tus propias pruebas, ejercicios de clase o prácticas de laboratorio. La prueba es gratuita, pero la herramienta tiene distintos costos y solo está disponible en inglés.
Questbase www.questbase.com	Es un sitio web que permite crear exámenes en línea y permite almacenar reactivos para compartirlos entre los usuarios. Es gratuito, pero únicamente está disponible en inglés.

EasyLMS https://www.onlineexambuilder.com/es/	Es una plataforma gratuita que permite diseñar y almacenar todo tipo de exámenes, también es posible compartirlos y aplicarlos en línea.
Exam.net https://exam.net/	Sitio web para crear exámenes en línea, para aplicarlos en clase con los estudiantes por medio de dispositivos electrónicos de manera segura, los exámenes cuentan con candados de seguridad. Además, se puede utilizar en 8 idiomas. Es gratuito

REFERENCIAS

Abedalqader, O., Hisham Bani, K., Abedalqader, B. & Hassan Bani, H. (2015). The Level of Test–Wiseness for the Students of Arts and Science Faculty at Sharourah and its Relationship with Some Variables. *Journal of Education and Practice*, *6*(29), 102-113. https://files.eric.ed.gov/fulltext/EJ1081246.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2018). *Estándares para pruebas educativas y psicológicas*. https://www.testingstandards.net/uploads/7/6/6/4/76643089/9780935302745_web.pdf

Burruezo Ordóñez, M., Cortés de las Heras, J., Martínez Soriano, V y Moreno Agud, A. P. (2014). *Diseño de Pruebas Objetivas. Cuadernos de instrumentos de evaluación #01*. https://mestreacasa.gva.es/c/document_library/get_file?folderId=500010855702&name=DLFE-522458.pdf

Canal IRCOficial. (2020, 7 de mayo). *Perfil de referencia*. [Video]. YouTube. https://www.youtube.com/watch?v=aTA_YJOgO8g

Canal IRCOficial. (2020, 7 de mayo). *Tabla de especificaciones*. [Video]. YouTube. https://www.youtube.com/watch?v=o5NXNnc5x54

Centeno Brambila, D. A. y Lira Obando, A. (2015). Sistema de evaluaciones en línea como herramienta para los niveles de educación media superior. *Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, *6*(11), 67-91. https://www.ride.org.mx/index.php/RIDE/article/view/119/533

Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia (2020). *Guía 2020 para preparar el examen de selección para ingresar a la licenciatura. Área de las Ciencias Biológicas, Químicas y de la Salud*. Universidad Nacional Autónoma de México.

Dlab, M.H., Katic, M.A., & Candrlic, S. (2015). *Ensuring formative assessment in e-course with online tests. 2015 10th International Conference on Computer Science & Education (ICCSE)*, 322-327. https://ieeexplore.ieee.org/document/7250264

Downing, S. M. & Haladyna, T. M. (2006). *Handbook of test development*. https://fatihegitim.files.wordpress.com/2014/03/hndb-t-devt.pdf

González, M., Marco, E. y Medina, T. (2020). Informe de iniciativas y herramientas de evaluación online universitaria en el contexto del Covid-19. https://www.usal.es/files/Informe_modelos_evaluacion_Gabinete_ministro_universidades.pdf

Haladyna, T. M, Downing, S. M. & Rodríguez, M. C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, *15*(3), 309-333. https://doi.org/10.1207/S15324818AME1503_5

Martínez González, A., Manzano Patiño, A., García Minjares, M., Herrera Penilla, C., Buzo Casanova, E. y Sánchez Mendiola, M. (2020). Perfil del estudiante con éxito académico en las licenciaturas del área de las Ciencias Biológicas, Químicas y de la Salud. *Revista de la Educación Superior*, *49*(193), 129-152. http://resu.anuies.mx/ojs/index.php/resu/article/view/1029/419

Martínez González, A., Sánchez Mendiola, M., Manzano Patiño, A., García Minjares, M., Herrera Penilla, C., y Buzo Casanova, E. (2018). Grado de conocimientos de los estudiantes al ingreso a la licenciatura y su asociación con el desempeño escolar y la eficiencia terminal. Modelo multivariado. *Revista de la Educación Superior*, *47*(188), 57-85. http://resu.anuies.mx/ojs/index.php/resu/article/view/508/269

Martínez, A., Herrera C., García, M., Ortega, E., Buzo, E. y Soto, G. (2021). *Evaluación para el aprendizaje: enfoque cuantitativo.* [MOOC] https://mooc.es/course/evaluacion-para-el-aprendizaje-enfoque-cualitativo/

Morales Vallejo, P. (2006). *Las pruebas objetivas: normas, modalidades y cuestiones discutidas.* https://cupdf.com/document/las-pruebas-objetivas-normas-modalidades-y-5o-las-preguntas-con-varias-respuestas.html?page=1

Phelps, R. P. (2017). The "Teaching to the Test" Family of Fallacies. *Revista Iberoamericana de Evaluación Educativa*, *10*(1), 33-49. https://doi.org/10.15366/riee2017.10.1.002

Rivera Jiménez, J., Flores Hernández, Alpuche Hernández, A. y Martínez González, A. (2016). Evaluación de reactivos de opción múltiple en medicina. Evidencia de validez de un instrumento. *Revista de Investigación en Educación Médica*, *6*(21), 8-15. https://doi.org/10.1016/j.riem.2016.04.005

UNAM-CUAIEED. (2020). *Evaluación del y para el aprendizaje a distancia: Recomendaciones para docentes de educación media superior y superior*. https://www.dee.cuaieed.unam.mx/index.php/evaluacion-del-y-para-el-aprendizaje-a-distancia-recomendaciones-para-docentes-de-educacion-media-superior-y-superior/

UNESCO (2020). *Exámenes y evaluaciones durante la crisis del COVID-19: prioridad a la equidad.* https://es.unesco.org/news/examenes-y-evaluaciones-durante-crisis-del-covid-19-prioridad-equidad

Universidad Autónoma de Chihuahua. (2012). *Reglamento para la aplicación de exámenes en línea.* http://www.fm.uach.mx/portal/2012/05/30/Reglamento%20de%20Examenes%20en%20Linea.pdf

Universidades Públicas de Castilla y León. (2020). *Guía de recomendaciones para la evaluación online en las Universidades Públicas de Castilla y León*. https://repositorio.grial.eu/handle/grial/2013

Capítulo 13

QUIZ

Sandra Paola Rodríguez Castillo, Careli Johana Herrera Penilla

"Sería una pena que después de todo esto volvamos a tener un sistema educativo como el que teníamos antes. Tenemos que salir de esto con sistemas educativos innovadores, robustecidos, distintos, que respondan mejor al contexto, que se adecuen a esta nueva realidad."
María Brown, Ministra de Educación de Ecuador

INTRODUCCIÓN

Ante las experiencias de aprendizaje remoto derivadas por la pandemia ocasionada por el virus SARS-CoV2 (COVID-19), las instituciones educativas han utilizado diferentes plataformas digitales para desarrollar de la mejor manera posible el proceso de enseñanza y aprendizaje, así como la evaluación asociada a este. En la búsqueda de instrumentos para valorar el aprendizaje de los estudiantes en el contexto actual, las tecnologías de la información y comunicación (TIC) resultan de gran ayuda, en tanto facilitan herramientas y recursos didácticos innovadores que pueden incorporarse al espacio educativo con el fin de analizar el progreso del alumnado.

En este capítulo se presenta el *quiz* como una forma sencilla de evaluar conocimientos o habilidades específicas a través de preguntas rápidas. El objetivo es ofrecer una guía conceptual y procedimental para la implementación de este instrumento, tanto en el contexto de la educación presencial, como a distancia y mixta, a más de orientar en el uso de sus resultados en aras de la mejora del proceso de enseñanza-aprendizaje.

Dada la gran variedad de sitios web y plataformas para construir este instrumento, es que se considera importante el contenido de este capítulo, a fin de que la elaboración y uso del *quiz* sean pertinentes y adecuados a las necesidades y particularidades de cada contexto educativo.

La información se encuentra organizada en diferentes apartados, en el primero se explica qué es, cuáles son sus ventajas y limitaciones; en el segundo se detallan los pasos para diseñarlo; en el tercero se hacen recomendaciones para su aplicación; posteriormente, con base en la investigación reciente, se señalan algunas características y utilidad del *quiz* en la educación y evaluación a distancia. Las siguientes secciones dan cuenta del análisis y utilidad de los resultados obtenidos a través de este instrumento; así como una serie de conclusiones y recomendaciones.

Como apoyo para el lector, se incluye una guía rápida para el uso de la plataforma *Kahoot!,* videos, ejercicios, además de recursos en línea que pueden utilizarse para desarrollar y guardar los quizzes.

Esperamos que este capítulo sea provechoso para enriquecer la enseñanza, el aprendizaje y la evaluación, como una triada perfecta para lograr el éxito dentro del ejercicio docente.

¿QUÉ ES?

El *quiz* o cuestionario corto[1], es un instrumento de evaluación compuesto por un conjunto pequeño de reactivos correctamente estructurados, generalmente, se ofrece un tiempo breve para su resolución y tiene como fin valorar conocimientos o habilidades específicas de quien lo responde.

Algunas de las características particulares de este instrumento y que lo hacen diferente respecto a un examen son:

- Frecuentemente es utilizado de manera lúdica entre docentes y estudiantes (Martínez, 2017). Por ello, es necesario que el número de reactivos que lo conformen y el tiempo destinado para resolverlo no sean excesivos; se recomienda que sean alrededor de 15 ítems y considerar de 15 a 20 minutos para su resolución.
- Dado que el tiempo de respuesta debe ser breve, se recomienda no utilizar reactivos de respuesta abierta o en los cuales el estudiante deba emplear mucho tiempo para su resolución, pueden ser reactivos de opción múltiple, falso-verdadero y/o de completamiento de espacios en blanco (Cohen y Sasson, 2016).

Ventajas

- Es un instrumento que se puede generar y aplicar tanto en entornos de la enseñanza presencial, como a distancia o mixta.
- Los estudiantes pueden responder desde sus teléfonos, tabletas o computadoras personales.
- Actualmente existe un amplio número de plataformas que permiten gestionar su elaboración, aplicación, resguardo y la obtención inmediata de resultados, por ejemplo, *Moodle, Google Classroom, Socrative,* entre otros.
- No se requiere mucho tiempo para su elaboración y dado que los resultados se pueden obtener de manera rápida y eficaz, es el instrumento idóneo para realizar ejercicios de retroalimentación inmediata, fomentando la evaluación formativa entre docente y alumnos o entre pares (Cohen y Sasson, 2016).
- Permite evaluar conocimientos y habilidades de razonamiento o solución de problemas.

---

1 Existen registros de la palabra quiz con el significado de "persona excéntrica", acto o acción de interrogar, específicamente una breve prueba oral o escrita (Merriam-Webster, s.f.).

- Ayuda a sintetizar el contenido de la asignatura o tema en cuestión.
- Proporciona a los estudiantes la oportunidad de autoevaluar su aprendizaje, es un instrumento útil para monitorear sesión con sesión las fortalezas y/o áreas de mejora a lo largo del curso.
- Promueve la finalización de las tareas porque los estudiantes saben que tendrán que usar el material para responder el cuestionario.
- Aumenta la motivación del alumno y promueve la participación en el proceso de aprendizaje.

Limitaciones

- Al ser un instrumento breve, únicamente se pueden evaluar contenidos demasiado específicos o poca cantidad de contenidos.
- Si no se aplica de manera adecuada, debido a la asignación de un tiempo determinado para su resolución, puede generar nerviosismo en los estudiantes, lo cual puede intervenir en los resultados.
- La falta de accesibilidad es un asunto por considerar. Si su elaboración y gestión es a través de un *software* o plataforma, todos los involucrados en la evaluación, deben contar con el equipo necesario para visualizarlo y responderlo.
- En ocasiones, un buen resultado puede ser más una cuestión de habilidades para responder el *quiz* que de comprensión del contenido (Conrad y Openo, 2018).

¿CÓMO LO DISEÑO?

Al delinear un instrumento de evaluación, pueden surgir distintas preguntas, por ejemplo, ¿cuál es el más apropiado?, ¿en qué momento del proceso de enseñanza y aprendizaje es pertinente aplicarlo?, ¿cuál es su finalidad? y ¿cómo ayudará a los estudiantes? Por consiguiente, el objetivo de esta sección es brindar una guía para la planeación y el diseño de un *quiz* de manera tradicional, es decir, en papel y lápiz, o a través de alguna plataforma que, como se verá más adelante, existen diferentes recursos en línea para su elaboración y almacenamiento.

Planeación

1) Defina el tipo de evaluación que va a realizar por medio del *quiz*, puede ser diagnóstica, formativa, sumativa, evaluación entre pares, autoevaluación, etc.
2) Pregúntese si echará mano del instrumento para evaluar a los estudiantes o serán ellos quienes se evalúen entre sí. Tal vez, ¿lo pueda implementar de manera grupal para alguna actividad?
3) Precise el tema o contenido que será evaluado por medio del *quiz*. Es deseable que únicamente se valore un tema o contenido.
4) Delimite los puntos sobresalientes o indispensables del tema o contenido que desea evaluar; recuerde que sobre esto versarán los reactivos.

5) Establezca el tipo de reactivos que contendrá el instrumento, recuerde que lo recomendable es que sean de opción múltiple, falso-verdadero y completamiento de espacios en blanco.
6) Elabore los reactivos, recuerde que lo recomendable es aplicar un máximo de 15 en aproximadamente 20 minutos.
7) Asigne un puntaje a cada uno de los reactivos elaborados, este puede variar dependiendo de la dificultad del reactivo o bien, puede ser el mismo para cada uno de ellos.
8) Integre y resguarde la cadena de respuestas correctas del instrumento completo, esto con la finalidad de agilizar el proceso de obtención de calificación.

Diseño

1) Defina el título del *quiz*. Se recomienda que el título sea llamativo, corto y no tan formal, a fin de que no pierda su naturaleza lúdica.
2) Coloque un espacio para anotar los datos del o los usuarios, así como datos de la asignatura, grupo, tema o actividad que se evalúa por medio del instrumento, fecha, nombre del profesor, etcétera, esto es de utilidad para documentar los procesos de evaluación que se realizan por medio del *quiz* y contar con evidencia de la evaluación.
3) Asiente las instrucciones completas para resolver el cuestionario, entre más completas estén será más fácil su aplicación. Indique si se debe subrayar, tachar o circular la respuesta correcta, el tiempo del que se dispone para resolverlo, el valor de cada reactivo y qué se debe hacer con el instrumento una vez resuelto en su totalidad.
4) Integre los reactivos que conformarán el instrumento; revisé que estén en el orden deseado y completos.
5) Coloque un espacio para anotar el puntaje total obtenido por el o los estudiantes al finalizar el *quiz*.
6) Realice una revisión final del instrumento completo, esto le permitirá asegurar que el *quiz* está completo y correctamente ensamblado.

Considere que:

- Es recomendable tener este instrumento primero a manera de borrador para poder hacer modificaciones en caso de encontrar errores ortográficos o de redacción una vez que se haya elaborado.
- La redacción de los reactivos debe realizarse empleando un lenguaje simple, claro y preciso, además, cada uno de ellos debe contener la información completa para contestarlo.
- Si se incluye retroalimentación para las respuestas, esta deberá ser en un tono positivo al tiempo que permita identificar fortalezas y áreas de mejora en el manejo de los contenidos.
- El nivel escolar, la edad y el capital cultural de los estudiantes son factores importantes.
- No se recomienda utilizar reactivos de respuesta abierta, pero en caso de hacerlo, es necesario que la respuesta sea de máximo cinco palabras y que la respuesta sea unívoca y precisa, esto evitará caer en la subjetividad y entorpecer el proceso de obtención de la calificación.

- En caso de utilizar reactivos de opción múltiple, se sugiere seguir todas las recomendaciones del capítulo titulado "Examen objetivo" de este libro.
- Si se incorporan esquemas, mapas, gráficas o imágenes a alguno o varios de los reactivos, es muy importante cerciorarse de que toda la información contenida en ellos sea visible y esté completa.

¿CÓMO LO APLICO?

Con el fin de utilizar el *quiz* como una forma de involucrar a los estudiantes en una aprendizaje activo y significativo, en el que se les anime a investigar, reflexionar y discutir los contenidos a través de preguntas teóricas y prácticas, a continuación, se enlistan una serie de puntos que orientan en el uso de este instrumento.

- La aplicación puede realizarse antes de iniciar la clase, para verificar o reforzar el aprendizaje de la sesión anterior; también puede llevarse a cabo al término de la revisión de una unidad o tema a fin de sintetizar la información y resaltar los puntos más importantes.
- Puede utilizarse para verificar que los estudiantes hayan realizado una lectura previa del tema que se revisará en clase.
- La aplicación de este instrumento puede realizarse de manera lúdica, pues se pueden organizar juegos y competencias de habilidades y conocimientos por medio de él. Emplear juegos de preguntas basados en aplicaciones móviles puede mejorar el rendimiento de los estudiantes, la interacción, la competencia, además resultan visualmente atractivos y fáciles de usar (Wilkinson et al., 2019).
- Es fundamental dejar claras las instrucciones y el tiempo justo para responder el *quiz*.
- Durante la aplicación, es deseable supervisar a los estudiantes y monitorear las actividades que se realizan, ya que como el tiempo es limitado, es preciso dar solución a las dudas que puedan surgir de manera inmediata.
- Si se organiza algún tipo de concurso basado en el *quiz*, se pueden establecer categorías o temáticas, por ejemplo, en Matemáticas se encontrarán "Conocimientos teóricos de las matemáticas" y "Solución de problemas". De esta forma, se estarán evaluando tanto aspectos teóricos como prácticos.
- En caso de que se quieran evaluar habilidades, es fundamental considerar el espacio requerido para realizarlo.
- Al concluir la aplicación se pueden o no recoger los quizzes, debido a que el análisis de los resultados puede ser de manera grupal o a través de una evaluación por pares.

EL QUIZ EN LA EVALUACIÓN A DISTANCIA

En el proceso de enseñanza y aprendizaje a distancia, los alumnos estudian de forma independiente el contenido educativo en un modo interactivo durante una cierta parte del tiempo; se someten a pruebas continuas; realizan trabajos bajo la guía del docente e interac-

túan con sus compañeros de manera virtual. Bajo este panorama, uno de los desafíos como profesor es implementar metodologías innovadoras para mantener la motivación, el compromiso y la concentración de los estudiantes, al tiempo que se abordan y evalúan temas variados en un tiempo limitado. A continuación, se presentan diferentes estudios que exploran el uso potencial del *quiz* como un instrumento práctico y fácil de usar, el cual impulsa predominantemente la evaluación formativa, como un circuito de retroalimentación en el que docentes y alumnos juegan roles activos, distintos pero complementarios para permitir el aprendizaje (Rezaei, 2015; Jones et al., 2018).

El *quiz*, en entornos de educación a distancia y mixtos existentes previos a la pandemia por COVID-19, ya contaba con algunas plataformas para su elaboración, uso y resguardo, de esto dan cuenta Cohen y Sasson (2016), quienes analizaron el uso de este instrumento en la plataforma *Moodle* en un entorno de educación mixta en la asignatura de Física. Entre los hallazgos se hace especial énfasis en que el uso de este instrumento es óptimo para realizar una evaluación predominantemente formativa, dado que es posible colocar una retroalimentación inmediata al acierto y error en cada uno de los reactivos que conforman el instrumento, o bien, si se compone de reactivos de opción múltiple, a cada una de las opciones de respuesta. Además, señalan que, de 120 estudiantes, el 76% manifestó que prefieren resolver un *quiz* en línea que resolver exámenes escritos y de 112 estudiantes, el 45% prefiere tener tres intentos para resolver cada reactivo.

La información que reporta este trabajo resulta de utilidad bajo el contexto actual en el cual se encuentra la educación a distancia y mixta, pues el número de docentes que han optado por utilizar este instrumento como una herramienta de evaluación formativa va en aumento y muchas de las plataformas en las cuales se trabajan las sesiones de clase –*Moodle, Google Classroom, Socrative, Zoom,* entre otras– ya han incorporado las funciones para elaborar, almacenar, compartir, aplicar y analizar los resultados de un *quiz*.

En el caso de *Socrative*, plataforma de uso frecuente en educación, a través de la cual los alumnos responden cuestiones propuestas por el profesor, mientras que este último sigue el progreso individual y grupal del curso en tiempo real, los resultados muestran que los estudiantes se adaptan fácilmente a este tipo de aplicaciones, se muestran propicios a responder las preguntas además de existir diferencias significativas en los puntajes de los exámenes finales de aquellos que utilizaron la herramienta (Green, 2016; Frías et al., 2016). *(Véase el capítulo* 7 *Aprendizaje potenciado por evaluación).*

Para conocer acerca de la experiencia de evaluación del proceso de enseñanza y aprendizaje mediante el uso de *Socrative* lo invitamos a ver el video titulado Evaluación del proceso de enseñanza-aprendizaje mediante el uso de Socrative | UOC (Universitat Oberta de Catalunya, 2020).

Kahoot!, Quizizz, Plickers, Flipquiz, Quizlet, Cerebriti, QuizStar, Pregunta2*, Online Quiz Creator* y *Trivinet* son otros de los sitios web, populares por la aplicación de quizzes, mismos que se pueden resolver desde una tableta electrónica o teléfono móvil. Particularmente, *Kahoot!* es una de las herramientas que mayor notoriedad ha tenido para el trabajo formativo en clase, es una plataforma que se presta para realizar actividades lúdicas, en este sentido,

se ha demostrado que la gamificación y los sistemas de respuesta interactiva benefician a las poblaciones de estudiantes en todas las materias, geografías, niveles de grado, edades, género y etnias (Martínez, 2017; Kirzner et al., 2021). Investigaciones señalan que su uso incrementa el aprendizaje de los alumnos, lo cual se refleja en los resultados obtenidos en las evaluaciones (Morales y Orgilés, 2019; Guzmán et al., 2018; Ismail y Mohammad, 2017; Marín et al., 2016).

Como parte de una revisión de literatura, Wang y Tahir (2020) encontraron que *Kahoot!* puede tener un efecto positivo en el aprendizaje para diversos contextos y dominios, en comparación con la enseñanza tradicional y otras herramientas; mejora la interacción maestro-alumno y entre alumnos; fomenta la participación, incluso de aquellos estudiantes que normalmente no hablan en clase; finalmente, notaron que los desafíos principales se relacionaron con problemas técnicos, preguntas y respuestas difíciles de leer en la pantallas, así como presión por el tiempo y miedo a perder.

Figura 1. Logotipo de la herramienta ***Kahoot!***



¿CÓMO DISEÑO EL QUIZ A DISTANCIA?

El diseño de un *quiz* en el contexto de la educación a distancia está totalmente ligado a la aplicación que se decida utilizar, por ello, es recomendable experimentar con varias plataformas, a fin de seleccionar la que permita cubrir las necesidades específicas del proceso de enseñanza y aprendizaje.

En este apartado, se muestra el diseño del *quiz* desde *Kahoot!* El primer requisito para trabajar con esta aplicación es contar con un correo electrónico de **Gmail**, si no se cuenta con ello, es necesario generarlo. Es recomendable que a manera de ejercicio interactivo se responda el **Play demo game** dentro de la plataforma, pero para ello deben seguirse los pasos que se muestran en el tutorial Cómo CREAR un KAHOOT! (TUTORIAL 2021)[2], en el que se ilustra la creación de un *quiz* en pocos minutos, por favor, antes de continuar leyendo este capítulo mire el tutorial completo.

Ahora que ya ha visto el tutorial, apoyándose en él y en los siguientes pasos, diseñe su primer *quiz* en **Kahoot!**

[2] En este tutorial puede haber algunas variaciones respecto a actualizaciones en la plataforma de Kahoot!, pero en esencia es el mismo procedimiento.

1) Ingrese a su buscador favorito (le sugerimos que sea Google) y coloque la palabra **Kahoot!**
2) Dé clic en la opción **https://kahoot.it** que es la página para jugar, y luego dé clic sobre **kahoot.com** que se encuentra en la parte central inferior de la página.
3) Dé clic en las palabras **Log in** que se encuentran en la parte superior derecha de la página.
4) Dé clic en las palabras **Log in with Google**.
5) Antes de comenzar a crear su primer *quiz*, le pedimos por favor que realice el **ejercicio interactivo** de la plataforma que dice **Play demo game**, este consta de 9 reactivos de preguntas de cultura general y le permitirá observar una pequeña demostración del funcionamiento de la aplicación de un *quiz* en la plataforma, después de realizar este ejercicio interactivo, comience a crear su *quiz* con el paso número 6 que es el siguiente.
6) Dé clic en el botón **Create kahoot**.
7) Dé clic en el botón **Create**.
8) En la parte superior izquierda de la página, en donde dice **Enter kahoot title…** dé clic y se abrirá una ventana en la cual debe colocar todos los datos que la plataforma va a solicitar para su *quiz*, como son:

    - Título (Title): Puede ser el que usted quiera, recuerde que de preferencia debe ser corto.
    - Descripción (Description): Esta es opcional y únicamente sirve para que otros colegas puedan saber de qué trata su *quiz*, puede poner ahí la asignatura, el tema y el objetivo del instrumento, por ejemplo: realizar una evaluación diagnóstica del tema "Las partes de la célula" para la asignatura de "Biología I".
    - Portada (Cover image): Aquí puede utilizar una imagen relacionada con el tema que ya tenga guardada en su computadora o bien puede utilizar alguna de la librería de imágenes de *Kahoot!*
    - Lenguaje (Language): Colóquelo en **español** o si es un *quiz* en otro idioma, puede configurarlo de la manera en la cual lo necesite.
    - Visibilidad (Visibility): Puede configurarlo para verlo únicamente usted o bien que sea público.
    - Video introductorio al *quiz* (Lobby video): Esto es opcional, puede copiar el link de algún video de youtube en ese espacio para que antes de iniciar el *quiz* se reproduzca ese video.
    - Música durante el *quiz* (Lobby music): Seleccione la música que se reproducirá al momento en el cual los estudiantes estén respondiendo el *quiz*, hay varias melodías que ofrece *Kahoot!*, en este caso, no puede subir música propia.

9) Ahora dé clic en **Done**.
10) Se abrirá un espacio para que comience a cargar los reactivos de su *quiz* uno por uno, debe dar clic en el lado izquierdo de la página en el botón **Add question** y luego en la opción **Quiz**.

11) Dé clic en el espacio en donde dice **Click to start typing your question** y escriba la base de su reactivo, puede anexar en la base una imagen de la librería o bien una imagen que ya tenga en su computadora, también puede añadir videos cortos o gifs.
12) Después de la base se debe ir añadiendo las opciones de respuesta, que en el caso de *Kahoot!* son mínimo dos, máximo cuatro.
13) Una vez cargadas las respuestas debe seleccionar la que es correcta, recuerde que debe ser únicamente una.
14) Después establezca el tiempo con el cual contarán los estudiantes para elegir su respuesta, en *Kahoot!* va de 5 a 240 segundos.
15) Cuando haya terminado de escribir el primer reactivo, nuevamente dé clic en **Add question** y **Quiz** y repita de los pasos 10 al 13.
16) Cuando haya terminado de cargar todos sus reactivos completos, dé clic en el botón **Done**.
17) A continuación, dé clic en el botón **Play** para seleccionar cómo es que desea aplicarlo.

A continuación, se verá cómo aplicar el *quiz* elaborado en la plataforma *Kahoot!*

¿CÓMO LO INSTRUMENTO?

Para familiarizarse con la aplicación del *quiz* que elaboró en *Kahoot!,* lo invitamos a realizar lo siguiente:

1) Seleccione cómo es que quiere configurar el uso del *quiz*, puede ser jugar en vivo (*Host live*), esto es utilizarlo en una clase vía *Zoom* compartiendo pantalla y que sus alumnos seleccionen su respuesta, ya sea en un celular o su propia computadora. Si es un juego en vivo debe decidir si se realizará en equipos (*Team mode*) o bien individual (*Classic*), si se dispone de pocos dispositivos, ya sean celulares o computadoras, lo recomendable es la opción de *Team mode*, si cada alumno cuenta con su propio dispositivo puede ser la opción *Classic*.
2) Sea la opción que haya elegido, configure (siguiendo el tutorial) las opciones que se presentan para poder iniciar con la aplicación del *quiz*, como son:

    - Habilitar la opción de usuarios inapropiados (*Friendly nickname generator*): Si se habilita esta opción la plataforma no les permitirá a los alumnos registrarse en el *quiz* con un nombre ofensivo o, como su nombre lo indica, inapropiado. Le sugerimos habilitarla.
    - General (General): Este conjunto de configuraciones le permitirá que los reactivos se presenten de manera aleatoria al igual que las respuestas, no le sugerimos que sea de esta manera, lo recomendable es que los reactivos se presenten en el orden en el cual usted ha decidido elaborar el instrumento.

3) Una vez configurada la aplicación del *quiz*, la plataforma arrojará un **Game PIN**, es decir un número conformado por siete dígitos, este número debe compartirlo con los alumnos para que ellos ingresen a la plataforma (ya sea por celular o en una computadora) e ingresen el **Game PIN**.
4) Usted irá visualizando en la plataforma el registro de cada uno de sus alumnos al *quiz* y una vez que todos se hayan registrado, debe dar clic en **Start** para que las preguntas y respuestas se vayan desplegando en la pantalla y los alumnos puedan seleccionar la respuesta que consideren correcta.
5) Al final se mostrará el puntaje obtenido, ya sea por equipos o de manera individual y de esta forma podrá realizar una revisión de cada reactivo y las repuestas y hacer un ejercicio de retroalimentación a manera de evaluación formativa.

Kahoot! tiene varias opciones para poder aplicar un *quiz*, la que se describe en el tutorial es una de ellas, pero también puede compartirle a sus alumnos el **Game PIN** y programar la aplicación del *quiz* de manera asincrónica, es decir, que ellos lo respondan para algún ejercicio en el cual no estén teniendo una interacción directa ni con el docente ni con algún o algunos compañeros y posteriormente revisar sus resultados, pues entre sus ventajas está que la plataforma, además de almacenar varios quizzes, también guarda los resultados estadísticos de la manera en la cual los estudiantes responden a cada uno de los reactivos, esto permite detectar aquellos reactivos en los que hay errores más frecuentes y aquellos que no se les dificultan.

Además de *Kahoot!*, le sugerimos que revise algunas otras aplicaciones, plataformas o sitios web en los cuales se puede diseñar y aplicar un *quiz*; en la sección de “Recursos en línea” que está al final de este capítulo, se enlistan varias sugerencias.

¿CÓMO ANALIZO SUS RESULTADOS?

Es conveniente establecer de antemano la forma en que se analizarán los resultados del *quiz*. Por ejemplo, ¿se emplearán para una evaluación formativa? Si es así, ¿se indicará qué preguntas se respondieron correctamente?, ¿se dará la respuesta correcta a las preguntas contestadas incorrectamente y dirigirá al estudiante a la parte específica del contenido del curso en el que se podría encontrar la respuesta? Otro propósito formativo puede ser familiarizar a los estudiantes con los cuestionarios y las preguntas típicas, preparándolos para realizar el cuestionario final del curso. Del mismo modo, si el *quiz* se calificará, ¿el estudiante podría realizarlo más de una vez?, ¿se cuenta con un conjunto de preguntas parcialmente nuevas para armar otra versión?

Entonces, considere que:

- El análisis de resultados de este instrumento se basa, sobre todo, en puntajes ya establecidos, por ello, es importante que durante la realización del instrumento cada uno de los

reactivos tenga un valor asignado o se establezca la escala de valoración de cada uno de los reactivos.

- En el caso del manejo de los puntajes en las plataformas o sitios web como *Kahoot!*, hay ocasiones en que no es posible asignar diversos valores a cada reactivo, hay plataformas que, en cambio, sí lo permiten, esto tiene que ver directamente con la manera en la que están configuradas. Por ello, es importante explorar varias y decidir cuál es la más adecuada a cada necesidad.
- De acuerdo con el puntaje obtenido en el *quiz*, se pueden valorar los temas, contenidos o tareas que sea necesario reforzar al tiempo que se realiza una retroalimentación oportuna y pertinente.
- Al analizar los resultados es importante considerarlos de manera individual o bien realizar una revisión a nivel grupal, pues es posible que algunos reactivos obtengan bajos puntajes en más de dos o tres estudiantes, lo anterior sería un indicador de que es necesario repasar dicho tema o bien preguntar a los estudiantes si todos los reactivos aplicados son claros y entendibles, porque el hecho de obtener malos resultados en la aplicación de un *quiz*, no siempre significa que el aprendizaje de los estudiantes es deficiente, a veces tiene que ver con la manera en la cual están siendo planteados los reactivos, por ello, al analizar los resultados de un *quiz*, es importante dialogar con los estudiantes para detectar claramente las fortalezas y áreas de oportunidad, tanto de los estudiantes, como del instrumento en cuestión.

EJEMPLOS

1. Quiz: Conjugación de verbos

Nombre del alumno: ______________________________ Grupo: ________

Nombre del profesor: _____________________________ Fecha: ________

Asignatura: __________________________ Puntaje obtenido: __________

Instrucciones: A continuación, se presentan 15 verbos en tiempo infinitivo; en 5 minutos escríbelos en participio regular. Cada respuesta correcta equivale a 2 puntos. Valor total del *quiz*: 30 puntos.

Tabla 1

Comer _______________________________	Jugar ________________________________
Correr _______________________________	Descansar _____________________________
Aburrir ______________________________	Beber ________________________________
Oler ________________________________	Alimentar _____________________________
Fracasar _____________________________	Crecer _______________________________
Creer ________________________________	Sentir ________________________________
Esperar ______________________________	Abundar ______________________________
Establecer ____________________________	

Fuente: Elaboración propia

2. Quiz: Conocimientos en Humanidades

Nombre del alumno: _______________________________________ Grupo: _________

Nombre del profesor: ______________________________________ Fecha: _________

Asignatura: __________________________________ Puntaje obtenido: ____________

Instrucciones: A continuación, se presentan tres bloques de preguntas con las categorías de Geografía, Historia Universal y Filosofía. Cada pregunta equivale a dos puntos, responde las preguntas en 10 minutos. Valor total del *quiz*: 30 puntos.

Tabla 2.

GEOGRAFÍA				
El gran desierto de Gobi se ubica en	¿Cuál de estos países no tienen acceso al mar Mediterráneo?	¿Cuál es el río más largo del mundo?	¿Cuál es el país más poblado del mundo?	¿Cuántos volcanes siguen activos en el mundo?
A) Sudamérica	A) Chipre	A) Yangtsé	A) Rusia	A) 1,500
B) Asia	B) Serbia	B) Nilo	B) China	B) 1,000
C) América central	C) Egipto	C) Amarillo	C) India	C) 2,000
D) África	D) Eslovenia	D) Amazonas	D) Indonesia	D) 1,700
HISTORIA UNIVERSAL				
Dios egipcio del Sol	¿En qué año termina la Edad Media?	Año de inicio de la Primera Guerra Mundial	Libro escrito por Adolf Hitler en la cárcel	La Crisis de los Misiles sucedió en
A) Ra	A) 1496	A) 1910	A) Mi lucha	A) Bolivia
B) Asar	B) 1492	B) 1922	B) El prisionero	B) Perú
C) Anuket	C) 1688	C) 1914	C) Noche azul	C) Cuba
D) Amon	D) 1515	D) 1939	D) La lucha	D) Puerto Rico
FILOSOFÍA				
Aristóteles escribió la	Platón fue discípulo de	Hegel es al idealismo como Marx al	Objeto de estudio de la Ontología	La Crítica de la razón pura es una obra escrita por
A) Retórica	A) Sócrates	A) Empirismo	A) La nada	A) Hegel
B) Ilíada	B) Aristóteles	B) Materialismo	B) La ciencia	B) Spinoza
C) Ética nicomáquea	C) Parménides	C) Racionalismo	C) El ser	C) Kant
D) Divina comedia	D) Heráclito	D) Humanismo	D) El conocimiento	D) Nietzsche

Fuente: Elaboración propia

Por último, para visualizar algunos de los quizzes en línea, mostramos algunas actividades educativas gratuitas desde *Educaplay* y *ThatQuiz*, relacionadas con comprensión lectora y matemáticas, respectivamente.

CONCLUSIONES Y RECOMENDACIONES

- El *quiz* se ha convertido en un apoyo frecuente en las prácticas de evaluación docente debido a su flexibilidad, inmediatez en los resultados y practicidad para una retroalimentación oportuna.
- Puede ser elaborado y aplicado en formato impreso o bien en línea, lo importante de este instrumento es que sea breve y el tiempo para contestarlo no sea de más de 20 minutos.
- Para comenzar a elaborarlo es necesario establecer de manera específica el tema que se evaluará a través de este.
- La aplicación debe ser breve y rápida, se sugiere ofrecer retroalimentación a los estudiantes, a fin de reforzar los aprendizajes o bien hacer las correcciones o aclaraciones necesarias sobre el tema en cuestión.
- Se puede trabajar con este instrumento de manera individual a grupal y también puede ser utilizado como un seguimiento del proceso de aprendizaje de los estudiantes.
- El uso del *quiz* permite que los docentes preparen escenarios diversos para el momento en que se reúnan con sus alumnos, de forma que la clase, presencial o a distancia, pueda ser adaptada a las necesidades de los estudiantes, en tiempo real (Medeiros y Bessa, 2017).
- Si bien, muchos de los sitios web para la evaluación de contenidos a través de los quizzes son gratuitos, también se cuenta con versiones avanzadas o *premium*, las cuales implican una inversión o costo que no todos los centros educativos o usuarios pueden cubrir. Por tal motivo, se requiere que cada docente examine las distintas posibilidades que las plataformas ofrecen para el logro de los objetivos de enseñanza, aprendizaje y evaluación.

RECURSOS EN LÍNEA

Tabla 3

Recurso	Descripción
Pregunta2 https://pregunta2.com	Es un sitio web gratuito que permite crear y almacenar diversos quizzes y *test* tanto de personalidad como de conocimiento y aplicarlos en línea.
Socrative https://socrative.com	Es una herramienta interactiva que permite crear quizzes, *test*, exámenes y todo tipo de dinámicas de trabajo en línea. Existe una versión para ser utilizada por profesores y otra para estudiantes, además de ser una aplicación móvil.
Kahoot!! https://kahoot.it	Es una plataforma y aplicación móvil que permite crear, almacenar y aplicar quizzes en línea. Cuenta con una versión gratuita y diversas versiones de paga que tienen más herramientas.
Quizizz https://quizizz.com	Es una plataforma que permite crear quizzes en línea y es de acceso gratuito. En ella el docente genera las preguntas y le proporciona al alumnado la página web y el código del cuestionario para responder desde un ordenador o dispositivo móvil.
Plickers https://plickers.com	Es una plataforma que permite crear, almacenar y aplicar quizzes de manera gratuita.
Cerebriti https://cerebriti.com	Es un sitio web que cuenta con diversos quizzes ya creados y almacenados sobre diversas asignaturas como Matemáticas, Ciencias, Geografía, Historia, entre otras. También permite crear nuevos quizzes de manera gratuita.
Onlinequizcreator https://onlinequizcreator.com	Es un sitio web de acceso gratuito que permite crear quizzes y encuestas, así como almacenarlas y aplicarlas. Esta más enfocado en encuestas de mercado, pero también se pueden ensamblar quizzes de conocimiento.
QuizStar http://quizstar.4teachers.org	Es un sitio web de acceso gratuito que permite crear, almacenar y aplicar quizzes en línea, hay un espacio para el trabajo del docente y otro para el trabajo con los estudiantes.

REFERENCIAS

Ben, C. & Root, E. (2001). *Is this a trick question? A short quide to writing effective test questions*. Kansas Curriculum Center. https://www.k-state.edu/ksde/alp/resources/Handout-Module6.pdf

Cohen, D. & Sasson, I. (2016). Online quizzes in a virtual learning environment as a tool for formative assessment. *Journal of Technology and Science Education*, *6*(3), 188-208. https://www.redalyc.org/pdf/3311/331147308004.pdf

Conrad, D. & Openo, J. (2018). *Assessment strategies for online learning. Engagement and authenticity.* AU Press. https://www.aupress.ca/books/120279-assessment-strategies-for-online-learning/

Frías, M., Arce, C. y Flores-Morales, P. (2016). Uso de la plataforma socrative.com para alumnos de

Química General. *Educación Química, 27*(1), 59-66. https://www.sciencedirect.com/science/article/pii/S0187893X15000658

Green, A. (2016). Significant returns in engagement and performance with a free teaching app. *The Journal of Economic Education, 47*(1), 1-10. https://www.tandfonline.com/doi/abs/10.1080/00220485.2015.1106359?journalCode=vece20

Guzmán Duque, A., Mendoza Paredes, J. y Tavera Castillo, N. (2018). Kahoot!: Un mecanismo de innovación para la educación universitaria. En R. Roig-Vila (Ed.) *El compromiso académico y social a través de la investigación e innovación educativas en la enseñanza superior*, 633-340. https://rua.ua.es/dspace/bitstream/10045/87728/1/2018-El-compromiso-academico-social-63.pdf

Jones, J., Bradley, G. & Love, S. (2018). The quick quiz: An unplugged assessment for the math educator's toolbox. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 91*(2), 78-84. https://doi-org.pbidi.unam.mx:2443/10.1080/00098655.2017.1386003

Kirzner, R., Alter, T. & Hughes, C. (2021). Online quiz as exit ticket: Using technology to reinforce learning in face to face classes. *Journal of Teaching in Social Work, 41*(2), 151-171. https://doi-org.pbidi.unam.mx:2443/10.1080/08841233.2021.1898521

Marín Jiménez, A. E., Montejo Gámez, J. y Campaña Gómez, J. R. (2016). Una propuesta para el refuerzo de conceptos matemáticos a través de Kahoot! Para alumnos de 1º de ADE del Campus de Ceuta. *Revista del CIDUI*, 3. https://www.cidui.org/revistacidui/index.php/cidui/article/view/1060/1024

Martínez, G. (2017). Tecnologías y nuevas tendencias en educación: aprender jugando. El caso de Kahoot. *Opción*, *33*(83), 252-277. https://www.redalyc.org/pdf/310/31053772009.pdf

Medeiros, R. & Bessa, A. (2017). MiniTeste: uma ferramenta ágil para aplicação de avaliações personalizadas. *Renote Revista Novas Tecnologias na Educação, 15*(1), 1-10. https://seer.ufrgs.br/renote/article/view/75126/42565

Merriam-Webster. (s.f.). Quiz. En el *diccionario Merriam-Webster.com*. Obtenido el 4 de agosto de 2021 de https://www.merriam-webster.com/dictionary/quiz

Morales, A. y Orgilés, M. (2019). El uso de Kahoot como recurso de evaluación continua en el Grado en Psicología. En R. Roig-Vila (Ed.) *Investigación e innovación en la Enseñanza Superior. Nuevos contextos, nuevas ideas*, 332-342. https://rua.ua.es/dspace/bitstream/10045/98887/1/Investigacion-e-innovacion-en-la-ES_031.pdf

Ismail, MA-A. & Mohammad, JA-M. (2017). Kahoot: a promising tool for formative assessment in medical education. *Education in Medicine Journa,l 9*(2), 19-26. https://eduimed.usm.my/EIMJ20170902/EIMJ20170902_02.pdf

Ortega, E. (2021). *Evaluación para el aprendizaje: enfoque cuantitativo. Quiz* [MOOC] https://es.coursera.org/lecture/evaluacion-aprendizaje-enfoque-cuantitativo/quiz-97HNz

Rezaei, A. (2015). Frequent collaborative quiz taking and conceptual learning. *Active Learning in Higher Education, 16*(3), 187-196. https://journals.sagepub.com/doi/10.1177/1469787415589627

Tics Brothers. (2020, 5 de marzo). *Cómo CREAR un KAHOOT (TUTORIAL 2021)* [Video]. YouTube. https://www.youtube.com/watch?v=pANtMqNWBek&t=304s

Trejo González, H. (2019). Recursos tecnológicos para la integración de la gamificación en el aula. *Tecnología, Ciencia y Educación*, 13, 75-117. https://www.tecnologia-ciencia-educacion.com/index.php/TCE/article/view/285

Universitat Oberta de Catalunya. (2020, 10 de julio). *Evaluación del proceso de enseñanza-aprendizaje mediante el uso de Socrative | UOC* [Video]. YouTube. https://www.youtube.com/watch?v=3iN0vFnRDzY

Wang, A. & Tahir, R. (2020). The effect of using Kahoot! for learning – A literature review. *Computers & Education, 149*, 1-22. https://www.sciencedirect.com/science/article/pii/S0360131520300208

Wilkinson, K., Dafoulas, G., Garelick, H. & Huyck, C. (2019). Are quiz-games an effective revision tool in anatomical sciences for higher education and what do students think of them? *British Journal of Educational Technology, 51*(3), 761-777. https://dialnet.unirioja.es/servlet/articulo?codigo=7371095

Capítulo 14

LISTA DE COTEJO

Roxana Sierra Gonzalez, Karla Patricia Sosa Ramírez, Virginia González Garibay

"La lista de cotejo es una de las herramientas de productividad más poderosas jamás descubiertas."
Brian Tracy

INTRODUCCIÓN

El ámbito educativo, como el resto de las actividades humanas, no escapó al impacto de la pandemia por la COVID-19. La enseñanza a distancia fue la solución para continuar el esfuerzo educativo a pesar del confinamiento.

Esto implicó ajustes acelerados tanto del lado del cuerpo docente como de la población escolar, no solo para poner en práctica sus habilidades tecnológicas sino también para superar la incertidumbre y ansiedad ocasionadas por la pandemia.

En este contexto, la evaluación del aprendizaje también tuvo que ser ajustada para responder a las condiciones de educación remota.

En la modalidad de educación presencial usualmente se emplean los exámenes como medio predilecto para evaluar el aprendizaje del alumnado. Sin embargo, existen otras opciones que pueden ayudar al profesorado a valorar lo que el alumnado aprende en escenarios digitales.

Uno de ellos es la lista de cotejo, que es un instrumento fácil de construir y que se puede adaptar a distintos temas y contextos educativos.

La lista de cotejo permite recopilar datos cuantitativos de manera rápida y ayuda tanto a docentes como a estudiantes a clarificar los aspectos que serán considerados para medir el aprendizaje.

Es un instrumento que puede ser muy útil para evaluar ejecuciones, procesos y productos sencillos o complejos. Si bien hay un conjunto de pasos para construirlo, su contenido puede ser prácticamente infinito, dependiendo de los objetivos de aprendizaje.

Al utilizarlo en combinación con otros instrumentos aporta información que permite al profesorado juzgar distintas aristas del aprendizaje de sus estudiantes.

El objetivo de este capítulo es poner al alcance de las y los docentes un conjunto de recomendaciones que les ayuden en el diseño y el uso de listas de cotejo en la evaluación del y para el aprendizaje.

En el primer caso, la lista puede aportar información para decidir si la o el estudiante reúne las características deseables en una ejecución o producto determinado y asignarle una calificación, en el segundo caso permite tanto al profesorado como al alumnado identificar los aspectos que están fallando para definir acciones de mejora.

El contenido de este capítulo está organizado en ocho secciones. En la primera se explica qué es una lista de cotejo, en la segunda se detallan los pasos para diseñarla, en la tercera se sugiere cómo aplicarla, en la cuarta se exponen algunos aspectos de este instrumento en la evaluación a distancia, en la quinta se describe cómo analizar sus resultados, en la sexta se incluyen algunos ejemplos, en la séptima se presentan las conclusiones y recomendaciones y, en la octava, se sugieren recursos en línea.

Este capítulo es un recurso de autoestudio, reflexión y consulta que puede leerse o no en orden. Pretende ser un punto de partida para que cada docente explore las posibilidades que ofrecen las listas de cotejo, practique su diseño, reflexione sobre sus prácticas de evaluación e intercambie ideas con sus colegas.

¿QUÉ ES?

- La lista de cotejo es un instrumento que relaciona acciones sobre tareas específicas, organizadas de manera sistemática para valorar la presencia o ausencia de estas y asegurar su cumplimiento durante el proceso de aprendizaje.
- Se le conoce con otros nombres: lista de control, lista de comprobación, lista de verificación, lista de confrontación, lista de corroboración y *checklist* o *checking list*. Dado que es un recurso nemotécnico para marcar y recordar pasos, ayuda a tener presentes aspectos que se deben cumplir. Existen distintos tipos de listas de cotejo, desde las más elementales como las de compra para el supermercado o las de tareas pendientes hasta las que sintetizan realizaciones complejas como la conducción de un avión o la integración de protocolos de investigación.
- Usualmente se emplea una escala de respuesta dicotómica –sí/no, logrado/no logrado, presente/ausente, correcto/incorrecto, cumple/no cumple– o solo una casilla de verificación.
- En el ámbito educativo puede ser utilizada para evaluar conocimientos, destrezas o conductas. También ayuda al profesorado a valorar el logro estudiantil con objetividad.
- Una de sus ventajas es que es un instrumento fácil de construir, flexible y ampliamente aplicable en distintos escenarios de aprendizaje. Además, la lista de cotejo puede ser empleada tanto por el cuerpo docente como por la población escolar, ya que proporciona evidencias sobre el logro o desempeño estudiantil en situaciones específicas.
- Asimismo, permite obtener información de manera inmediata, lo que ayuda a definir el rumbo de acción a quienes participan en un proceso educativo, es decir, tomar decisiones

fundamentadas para mejorar el aprendizaje del alumnado.

- No obstante, también la lista de cotejo presenta algunas limitaciones, como no valorar el grado de cumplimiento de las actividades descritas y disminuir los detalles de la observación, por lo que su alto grado de generalidad puede conducir a sesgos.
- La lista de cotejo puede ser empleada con propósitos de diagnóstico, por ejemplo, para identificar las habilidades de lectura del alumnado al inicio de un curso; con propósitos formativos, cuando la o el docente quiere identificar qué conocimientos han logrado sobre un determinado tema para planear acciones de apoyo, o con propósitos sumativos, para hacer un recuento de lo aprendido al final de una unidad académica o de un ciclo escolar.

A continuación, se detallan los pasos para su diseño.

¿CÓMO LA DISEÑO?

Si bien las listas de cotejo son instrumentos simples, es necesario dedicar un tiempo a la planeación de su contenido y su diseño para asegurar que sirvan para el propósito para el que fueron creadas. Son muy valiosas cuando están bien diseñadas, revisadas y aplicadas. En esta sección se abordan esos temas.

Planeación

1) Seleccione el o los objetivos de aprendizaje que quiere valorar.
   ¿La lista de cotejo es el mejor instrumento para valorar el aprendizaje seleccionado?
2) Determine las actividades, tareas o aspectos de la lista de cotejo.
   Analice el o los objetivos seleccionados y con ello elija las actividades, tareas o aspectos que incluirá en la lista. Pueden ser planteados como preguntas o como afirmaciones.
3) Defina el tipo de evaluación que quiere realizar con la lista.
   Diagnóstica, formativa o sumativa.
4) Contemple el público al que se dirigirá la lista.
   Estudiantes, grupos de estudiantes, usted mismo o un grupo de docentes.
5) Determine el nivel de logro de las actividades.
   Cumplimiento de todas las tareas planteadas o solo algunas de ellas.
6) Defina los criterios para evaluar el desempeño por medio de la lista.
   Asigne un valor a cada actividad cumplida o al conjunto de actividades descritas en la lista.
7) Organice el listado de las actividades.
   Ordene las actividades, tareas o aspectos de acuerdo con su contenido. Si forman parte de un proceso, organícelas de manera secuencial. Si son similares, agrúpelas en secciones o categorías.

Notas:

- Las actividades deben ser claras, precisas y útiles para orientar la valoración del aprendizaje.
- Puede diseñar una versión en espejo de la misma lista, una para usted y otra para el alumnado.
- Las actividades pueden o no tener el mismo valor.
- Verifique la pertinencia, exhaustividad y unicidad de las actividades o aspectos de la lista.

Diseño

1) Defina el título de la lista de cotejo.
   Corto, sintético, sencillo y atractivo.
2) Incluya una sección de datos de identificación.
   Nombre de la asignatura, temática que apoya, datos de la o el docente, datos de la o el estudiante y fecha.
3) Redacte las instrucciones para su empleo.
   Explique con claridad y precisión las condiciones para el uso de la lista de cotejo.
4) Incorpore las actividades, tareas o aspectos.
   Organícelos en dos columnas: una donde se describa la actividad o tarea contemplada y otra que incluya un espacio para su verificación.
5) Considere un apartado para observaciones.
   En el cual sea posible expresar comentarios adicionales o donde el alumnado describa las razones por las que pudo o no completar la actividad definida. También se puede añadir al final de la lista.
6) Edite la versión final.
   Revise la lista y haga los ajustes para su edición, ya sea que se despliegue en una plataforma o aplicación o que se imprima para distribuirla con sus estudiantes.

Notas:

- La verificación puede ser en una o dos casillas.
- Si es el caso, destaque las actividades o aspectos de mayor importancia en la lista.
- Emplee distintas formas de presentación (colores, tipografías, imágenes). Los estudiantes aprecian las formas de diseño novedosas.

Consulte el video *Listas de Cotejo-¿Qué son y cómo elaborarlas?*
https://youtu.be/S6MnTr-1CPI (Pedagogía Digital, 2016).

Revisión

- Es recomendable que revise periódicamente la lista para asegurar su vigencia y pertinencia como instrumento de evaluación del y para el aprendizaje.

- Una vez aplicada, es muy útil que pida a las y los estudiantes comentarios o sugerencias sobre esta, puesto que son quienes la usan y su opinión es muy importante para mejorarla.
- También puede compartirla con sus colegas para que valoren la pertinencia, claridad y organización de las actividades, además del diseño.

¿CÓMO LA APLICO?

- Las listas de cotejo pueden ayudarle a evaluar el cumplimiento de tareas o actividades específicas en diferentes momentos del aprendizaje, por lo cual pueden aplicarse al inicio del curso, durante el mismo o al final.
- La evaluación puede ser individual –de cada estudiante– o colectiva –de un conjunto de estudiantes o de todo el grupo–. La lista de cotejo también puede ser respondida por usted o por un conjunto de docentes.
- Este instrumento se puede incorporar fácilmente a las acciones cotidianas de enseñanza. Por ejemplo, alumnado y profesorado pueden emplear las listas de cotejo para organizar mejor sus actividades y tareas pendientes, verificar el logro de un determinado número de actividades o la totalidad de estas, aprender una secuencia, valorar una tarea o llevar un registro individual del avance estudiantil y un registro global de los logros obtenidos por el grupo.
- Por la multifuncionalidad de la lista de cotejo, su aplicación puede hacerse tanto en espacios físicos (salón de clases, laboratorio, taller o en casa) como en espacios virtuales, esto depende de sus propósitos y de a quién vaya dirigida.
- Por ejemplo, puede emplear una lista para valorar la expresión oral, en este caso, un listado con enunciados que describan el desempeño ideal le permitirá observar a sus estudiantes y determinar su nivel de logro.
- La lista de cotejo se puede imprimir en papel para ser utilizada por usted o entregada a cada estudiante, pero también se puede elaborar una versión digital que se emplee desde un dispositivo electrónico –teléfono inteligente, tableta o computadora personal– o que se distribuya en un archivo por correo electrónico, incluso, se pueden emplear aplicaciones o páginas web diseñadas específicamente para ello (ver sección Recursos en línea).
- Al aplicar la lista de cotejo es necesario definir si esta será entregada antes, durante o al concluir una tarea. Por ejemplo, al inicio para que el alumnado se familiarice con un procedimiento, durante la ejecución de una actividad para que siga paso a paso determinadas instrucciones o al final de una tarea para que identifique sus aciertos o errores.
- También es importante determinar el tiempo que se empleará para su uso o el periodo que el alumnado tendrá para resolverla y devolverla.

Notas:

- ❍ Puede distribuirse para que las y los estudiantes la utilicen como referencia.
- ❍ Es útil para que usted verifique la planeación de actividades dentro y fuera de la clase, como recordatorio de diferentes tareas o para evaluar el cumplimiento de objetivos de las sesiones.

LA LISTA DE COTEJO EN LA EVALUACIÓN A DISTANCIA

¿Cómo la diseño a distancia?

- Dependiendo del formato en que usted planeé proporcionar al alumnado la lista de cotejo (en una plataforma electrónica, o en un archivo de *Word*, *Excel* o PDF, para enviarla por correo electrónico), será el tipo de herramienta que utilizará para diseñar este instrumento de evaluación.
- Si decide la primera vía, existen múltiples plataformas y aplicaciones en línea en las cuales puede diseñar y gestionar una lista de cotejo de manera rápida y sencilla, para realizar un seguimiento de las actividades (ver sección Recursos en línea).
- Para ello solo requiere acceder a una de estas herramientas, registrarse y comenzar a crear su lista. Puede visualizar en el espacio de trabajo las listas de cotejo que tenga pendientes, con el conjunto de actividades, tareas o aspectos que haya determinado para valorar su actividad de aprendizaje.
- Antes de diseñarla, es recomendable que disponga de una versión previa de la lista. Esto permitirá contar con un respaldo de información, si es que usted o el alumnado tienen dificultades durante su resolución en la plataforma elegida.
- En caso de que decida compartirla como documento de trabajo, enviándola por correo electrónico a sus estudiantes, o subiéndola como archivo a una plataforma educativa, defina con claridad las condiciones en que las y los estudiantes tendrán que utilizarla. Si requiere que la respondan en el mismo documento, es recomendable que la diseñe en *software* de uso extendido (*Word*, *Excel*, PDF).
- Se sugiere que, desde las instrucciones para su empleo, aclare a las y los estudiantes el significado que tendrá que marquen las casillas de verificación, para que sepan con claridad lo que deben hacer.
- Si tiene pensado estructurar su lista en categorías, tendrá que valorar las herramientas disponibles para editarla bajo ese formato. También la puede preparar en un documento de trabajo y distribuirla a sus estudiantes, aunque esto requerirá de inversión de tiempo para documentar los resultados y emitir las evaluaciones individuales necesarias.

¿Cómo la instrumento?

- Algunos de estos recursos son de uso personal, por lo que si diseña una lista de cotejo tendrá que descargarla en un formato que pueda compartir con el alumnado, ya sea que decida enviárselas por correo electrónico, alguna red social o compartírselas mediante una plataforma educativa (*Google Classroom*, *Moodle*, *Blackboard*).
- Otras herramientas permiten compartir las listas de cotejo, visualizarlas y editarlas en la misma plataforma, lo cual es ampliamente útil cuando la enseñanza se lleva a cabo de manera remota.
- Dado que las prácticas de evaluación a distancia no siempre son sincrónicas, es recomendable que las y los estudiantes puedan resolver la lista de cotejo en la herramienta

que usted ha utilizado. Esto permitiría planear tiempos de entrega específicos y llevar un control en tiempo real de lo que están resolviendo sus estudiantes.

- Es importante que mantenga contacto cercano con el alumnado para darle orientación y resolver sus dudas en relación con la lista de cotejo.
- Una vez que sus estudiantes le proporcionen las listas de cotejo resueltas, debe diseñar una estrategia para organizarlas y analizarlas; puede elaborar una hoja de cálculo donde registre sus evaluaciones para comparar resultados a nivel individual y grupal. Para retroalimentar al alumnado, puede expresar comentarios sobre su desempeño en la plataforma que haya empleado o en el documento de trabajo que proporcionó al grupo. Para simplificar la tarea, programe una sesión de trabajo virtual con sus estudiantes y proporcióneles un breve resumen del desempeño obtenido por el grupo, aunque complementándolo con la evaluación individual.

¿Cómo crear una lista de cotejo en *Todoist*?

- Esta herramienta permite crear listas de cotejo y organizarlas de acuerdo con su contenido. Cada actividad puede incluir subtareas y comentarios. También pueden programarse en tiempos específicos. Ingrese a la aplicación en línea www.todoist.com
- Oprima en "Comienza ahora".
- Regístrese en la plataforma, puede utilizar su cuenta de *Google, Facebook, Apple* o emplear su correo electrónico.
- Acceda con su usuario y contraseña. Visualice el área de trabajo y explore las diferentes opciones que tiene para diseñar una lista de cotejo.
- Para crearla, oprima el símbolo "+" del submenú "Proyectos", ubicado en una columna del lado izquierdo de la plataforma.
- Defina las características principales de su lista (nombre, etiqueta de color, si desea agregarla a favoritos) y oprima "Añadir". En la parte superior del área de trabajo aparecerá una opción para añadir tareas, oprima en esta, tantas veces como sea necesario, dependiendo del número de actividades, tareas o aspectos que desee evaluar con la lista de cotejo.
- A cada tarea le puede asignar una fecha de vencimiento, grado de prioridad y una breve descripción de sus características.
- Para visualizar sus detalles, oprima el botón izquierdo del ratón en cada una. Puede incluir subtareas, comentarios y verificar el historial de cada actividad.
- Además, es posible modificar el orden en que aparecen los elementos de la lista. Para ello, oprima el símbolo que se sitúa del lado izquierdo de la tarea "⠿" y desplácela a su nueva ubicación.
- Si posiciona el cursor al final de las tareas, podrá visualizar el menú que le permitirá ajustar sus características, duplicarlas, eliminarlas, cambiarlas de proyecto o compartirlas.
- También puede ajustar las características generales de la lista o proyecto, como su visualización y organización en el área de trabajo. Si lo desea, puede agrupar las tareas en categorías.

- Por último, distribuya la lista de cotejo. Seleccione la opción "Compartir", que se ubica en el menú principal de la lista e ingrese los correos electrónicos de sus estudiantes para que reciban la invitación para participar en el proyecto.
- Si tiene alguna duda adicional, puede acceder al menú de Ayuda que se encuentra en la esquina superior derecha de la plataforma, con guías de apoyo para aprender a utilizar *Todoist*.

Nota:

- Para conocer más detalles sobre esta plataforma consulte la sección de Recursos en línea al final del capítulo.

EJERCICIO DE AUTOEVALUACIÓN

Instrucciones: Marque con una X los elementos que cumplió al emplear la plataforma para diseñar su lista de cotejo, responda con sinceridad. Después sume los puntos que marcó y al final identifique cuál es la calificación que obtuvo en el ejercicio.

Tabla 1

☐	Exploró el área de trabajo y las distintas opciones para diseñar una lista de cotejo.
☐	Diseñó un proyecto, con al menos cinco tareas, actividades o aspectos.
☐	Redactó con claridad y precisión las tareas, actividades o aspectos.
☐	Definió las características de las tareas (fecha de vencimiento, grado de prioridad, descripción).
☐	Incorporó algunas subtareas o comentarios para las tareas, actividades o aspectos.
☐	Modificó el orden de aparición de los elementos de la lista.
☐	Ajustó las características particulares de una tarea, actividad o aspecto (la eliminó, duplicó o cambió de proyecto).
☐	Revisó las características generales de la lista (visualización, organización).
☐	Agrupó los componentes en categorías.
☐	Compartió la lista.
☐	Accedió al menú de Ayuda de la plataforma.

Puntuación	Desempeño
11	Excelente
10	Muy bueno
9	Bueno
8	Regular
7	Deficiente

Fuente: Elaboración propia

¿CÓMO LA ANALIZO?

- Los resultados de las listas de cotejo pueden ayudar a tomar decisiones sobre el avance de las y los estudiantes en el curso.
- El nivel de complejidad establecido para la lista de cotejo y su tipo de diseño ayudará a definir su forma de evaluación.
- Para el análisis de sus resultados es necesario tomar en cuenta si se trata de una evaluación a cargo de usted, en la cual tendrá que asignar un valor al desempeño de sus estudiantes; o de una autoevaluación de las y los estudiantes, quienes señalarán el cumplimiento de las actividades incluidas en la lista.
- En listas no secuenciales se puede asignar un valor a cada actividad cumplida y al final del ejercicio, se contempla el puntaje total obtenido por cada estudiante. En listas secuenciales el valor dependerá del cumplimiento de todas las actividades.
- Sin importar el tipo de lista elegida, usted tiene que contemplar los diferentes escenarios de resultados para valorar el nivel de desempeño de sus estudiantes.
- Dependiendo el tipo de evaluación y objetivos planteados, puede diseñar una escala que oriente la interpretación de los resultados.
- Por ejemplo, en una lista no secuencial de diez actividades, usted puede determinar que el cumplimiento de dos o menos actividades equivalga a un nivel de desempeño muy inadecuado; de tres a cinco, inadecuado; de seis a ocho, adecuado, y nueve o diez, muy adecuado.
- El análisis de resultados también dependerá de a quién o quiénes se dirige la lista (una o un estudiante, un conjunto de estudiantes o a todo el grupo).
- A nivel individual, usted puede comparar los resultados obtenidos por cada estudiante con el nivel esperado de conocimiento.
- A nivel grupal, puede analizar la tendencia o variación en el resultado global obtenido por el conjunto de estudiantes o todo el grupo para identificar las áreas que requieren de mejora y establecer directrices que fortalezcan el aprendizaje del alumnado.
- Como se señaló en la primera sección de este capítulo, la ausencia o presencia en el cumplimiento de las actividades que contiene la lista de cotejo es un indicador del desempeño esperado en el alumnado.
- A simple vista, usted puede apreciar en la lista dicha ausencia o presencia para obtener un resultado prácticamente inmediato.
- Si emplea herramientas o aplicaciones en línea, como por ejemplo *Google Classroom*, *Blackboard* o *Moodle* el análisis de los resultados es más sencillo, porque no solo puede consultar sino también almacenar los resultados individuales y grupales.

EJEMPLOS

1. Evaluación del desempeño en las presentaciones.

Instrucciones: Elija la opción que refleje el desempeño estudiantil en cada una de las características descritas.

Tabla 2

Sí	No	
☐	☐	Presenta la información con claridad.
☐	☐	Utiliza un tono de voz adecuado (claro y preciso).
☐	☐	Expone los contenidos con detalle.
☐	☐	Emplea un lenguaje apropiado.
☐	☐	Organiza la información de lo general a lo particular.
☐	☐	Se apoya de materiales o ejemplos para clarificar las ideas.
☐	☐	Relaciona el tema con los contenidos del curso.
☐	☐	Utiliza adecuadamente su tiempo.

Aspectos cumplidos	Desempeño	Calificación
8	Excelente	10
7	Muy bueno	9
6	Bueno	8
5	Regular	7
4	Deficiente	6
3 o menos	Insuficiente	5

Fuente: Elaboración propia.

2. Elaboración de un ensayo

Instrucciones: Marca con una X si tu trabajo cumple con los siguientes elementos. Al final se presenta un espacio de observaciones para explicar por qué no se cumplen.

Tabla 3

Nombre de la asignatura:	
Tema:	**Nombre de la o el docente:**
Nombre de la o el estudiante:	**Fecha:**

Introducción	☐	Introduje correctamente al tema.
	☐	Expliqué el argumento principal.
	☐	Definí el contexto del tema.
Desarrollo	☐	Describí con claridad el argumento principal.
	☐	Expuse de manera detallada las ideas.
	☐	Fundamenté los argumentos en literatura sobre el tema.
Conclusión	☐	Expuse ideas o argumentos a favor y en contra.
	☐	Establecí conexiones entre la información.
	☐	Presenté una reflexión final sobre el tema.
Estilo/Edición	☐	Presenté las ideas en forma lógica y clara.
	☐	Utilicé un lenguaje formal.
	☐	Redacté las ideas correctamente.
	☐	La gramática y ortografía del ensayo son correctas.
Referencias	☐	Me basé en evidencia confiable.
	☐	Incorporé las referencias al final del ensayo.
	☐	Cité las referencias en el mismo formato.
Observaciones		

Aspectos cumplidos	Desempeño	Calificación
16	Excelente	10
13-15	Bueno	9
10-12	Regular	8
7-9	Deficiente	7
6 o menos	Insuficiente	6

Fuente: Elaboración propia.

CONCLUSIONES Y REFLEXIONES FINALES

- La lista de cotejo es un instrumento que permite valorar la ausencia o presencia de acciones, tareas o aspectos para verificar su cumplimiento durante el proceso de aprendizaje.
- Es un instrumento versátil que puede ser adaptado a múltiples temas y contextos educativos con diferentes grados de complejidad.
- Permite recopilar datos cuantitativos de manera rápida, lo que proporciona referentes para valorar el aprendizaje del alumnado en un periodo corto de tiempo.
- Es útil para evaluar no solo al alumnado, sino también al profesorado o a un equipo docente. Puede diseñarse en forma de espejo para comparar la valoración de las diferentes personas involucradas.
- Sus beneficios comprenden no solo a la evaluación, sino también a la planeación y a la conducción del curso.
- La lista de cotejo puede ser empleada en diferentes momentos del aprendizaje.
- Por su multifuncionalidad, la lista puede aplicarse tanto en espacios físicos de aprendizaje (salones de clase, laboratorios, talleres y otros lugares) como en espacios virtuales.
- Puede diseñarse y administrarse a través de aplicaciones, herramientas, plataformas o páginas *web*.
- Es un instrumento que permite comparar los resultados de un gran número de participantes, por lo que, además de las valoraciones individuales, la o el docente dispondrá de los resultados de todo el grupo.
- Es recomendable que antes de construir una lista de cotejo se reflexione si este instrumento es el más idóneo para valorar el aprendizaje del alumnado y si proporcionará evidencias útiles para la evaluación dentro del curso.
- Es importante someter la lista de cotejo a una revisión periódica para asegurar su vigencia y pertinencia como instrumento de evaluación. En este sentido, es útil consultar a las y los usuarios finales, ya sean estudiantes o docentes, para contar con comentarios o sugerencias que ayuden a mejorarla.

RECURSOS EN LÍNEA

Tabla 4

Recurso	Descripción
Any do https://desktop.any.do/	Permite crear listas de cotejo y clasificarlas de acuerdo con categorías o etiquetas y programar diversas opciones de recordatorios. Pueden incluirse subtareas, notas o archivos adjuntos para cada una de las actividades.
Canva https://www.canva.com/es_us/crear/checklist/	Sirve para diseñar listas de cotejo, ampliamente personalizables, que pueden ser visualizadas en dispositivos electrónicos sin necesidad de instalar ninguna aplicación o programa, únicamente se requiere acceso a internet.
Evernote https://www.evernote.com/	Ayuda a crear listas de cotejo de manera sencilla. Se pueden asignar recordatorios o etiquetas a las tareas y compartir a través de las redes sociales. Las listas se pueden sincronizar con los dispositivos electrónicos de la o el usuario.
Google Keep https://keep.google.com/	Permite diseñar y gestionar listas de cotejo en un área de trabajo. Es posible añadir recordatorios a las tareas, personalizarlas con diversos colores, incluir imágenes y compartirlas. Solo se necesita ingresar con un correo electrónico.
Tick Tick https://ticktick.com	Ayuda a crear listas de cotejo y organizarlas según su fecha de creación u orden de prioridad. Cuenta con un menú que resume el avance de las actividades de la o el usuario en un periodo específico de tiempo. Se pueden exportar en distintos formatos.
To do https://todo.microsoft.com	Cuenta con múltiples posibilidades para crear listas de cotejo, organizar sus tareas, clasificarlas por colores y establecer avisos únicos o continuos. Se pueden descargar o imprimir sin necesidad de un pago adicional.
Todoist https://todoist.com/es/	Permite diseñar listas de cotejo de forma sencilla y clasificarlas de acuerdo con su contenido, y su orden de prioridad, así como dividir en subtareas y programarlas en tiempos específicos. Se pueden compartir con diversos colaboradores.
Venngage https://es.venngage.com	Ayuda a crear listas de cotejo a partir de plantillas preestablecidas o con un formato propio diseñado por la o el usuario. Ofrece múltiples tipografías, estilos y formatos para diseñar las listas y las tareas que involucra cada una.
Zenkit https://zenkit.com/en/todo/	Permite crear listas de cotejo con un diseño atractivo e interactivo. Dispone de diversas opciones para organizar las tareas. Se pueden establecer alertas para el cumplimiento de las actividades y filtrar por su contenido.

REFERENCIAS

Bichelmeyer, B. (2003). *Checklist for Formatting Checklist. Evaluation Checklist Project.* The Evaluation Center. Evaluation Checklists. Michigan, Estados Unidos: Western Michigan University. Recuperado de https://bit.ly/2CPGznf

British Columbia Institute of Technology. (2010). *Developing Checklists and Rating Scales.* Learning and Teaching Centre. Burnaby, Canadá. Recuperado de https://www.bcit.ca/files/ltc/pdf/ja_developchecklists.pdf

Edmunds, J. (2006). *How to Assess Student Performance in History: Going Beyond Multiple-Choice Tests.* Greensboro, Estados Unidos: Regional Educational Laboratory Southeast. Recuperado de: https://files.eric.ed.gov/fulltext/ED513873.pdf

Learner-Centered Initiatives. (2010). Checklist for Assessment to Produce Learning. Recuperado de http://lciltd.org/tools/Assessment/Checklist_AssessmentToProduceLearning.pdf

Marino, K. (2013). *How a Simple Checklist Can Improve Learning. TeachThought.* Recuperado de http://www.teachthought.com/pedagogy/simple-checklist-can-improve-learning/

Pedagogía Digital. (2016, 16 de marzo). *Listas de Cotejo - ¿Qué son y cómo elaborarlas?* [Video]. Youtube. https://youtu.be/S6MnTr-1CPI

Pérez, C. (2018). Uso de listas de cotejo. Una guía para el profesor. Universidad Tecnológica Metropolitana. Vicerrectoría Académica. Unidad de Mejoramiento Docente. Recuperado de https://vrac.utem.cl/wp-content/uploads/2018/10/manua.Lista_Cotejo-1.pdf

Reijers, H. A., Leopold, H. y Recker J. (2017). *Towards a Science of Checklists*. Recuperado de https://scholarspace.manoa.hawaii.edu/bitstream/10125/41859/paper0710.pdf

Rowlands, K. (2007). Check It Out! Using Checklists to Support Student Learning. *The English Journal, 96*(6), 61-66. Recuperado de http://www.jstor.org/stable/30046754

Scriven, M. (2007). *The logic and methodology of checklists.* The Evaluation Center. Evaluation Checklists. Michigan, Estados Unidos: Western Michigan University. Recuperado de https://www.semanticscholar.org/paper/THE-LOGIC-AND-METHODOLO-GY-OF-CHECK€LISTS-Scriven/263b77fec3e3bb7d3298399b4672b5ebf1c2c1ab

Stufflebeam, D.L. (2000). *Guidelines for developing evaluation checklists: The checklists development checklist* (CDC). The Evaluation Center. Evaluation Checklists. Michigan, Estados Unidos: Western Michigan University. Recuperado de https://wmich.edu/sites/default/files/attachments/u350/2014/guidelines_cdc.pdf

WIDA (Wisconsin, Delaware y Arkansas) Consortium. (2009). WIDA Focus on Formative Assessment. *Focus Bulletins, 1*(2), 1-6. Wisconsin, Estados Unidos: University of Wisconsin–Madison Wisconsin Center for Education Research. Recuperado de https://bit.ly/372kjo9

Capítulo 15

RÚBRICA

Roxana Sierra Gonzalez, Karla Patricia Sosa Ramírez, Virginia González Garibay

"El proceso mismo de crear una rúbrica me ayuda como educador a comprender mejor el 'desempeño' esperado."
Christopher R. Gareis

INTRODUCCIÓN

La evaluación del aprendizaje siempre ha sido un reto para el profesorado, a este hoy se suma el impacto de las condiciones ocasionadas por la pandemia de la COVID-19. Este hecho, si bien es un desafío también es una oportunidad para las comunidades académicas de incorporar nuevas herramientas para la educación en los escenarios digitales.

En los escenarios emergentes de educación a distancia, el profesorado ha tenido que reforzar su función como facilitador del aprendizaje de sus estudiantes a fin de que asuman un rol protagónico en la construcción de sus aprendizajes. Por otro lado, ha tenido que apoyarse en la evaluación continua del aprendizaje para proporcionarles retroalimentación y apoyo certero.

Sin importar la modalidad educativa, la evaluación del aprendizaje debe formar parte de la enseñanza. Para juzgar con suficiencia y pertinencia el aprendizaje, es necesario contar con evidencias que permitan documentarlo y analizarlo para ayudar a perfeccionarlo. La información que aporta la evaluación, además de ser útil para el profesorado, ayuda al alumnado a ser más autónomo y asumir la responsabilidad de su aprendizaje.

La evaluación del y para el aprendizaje es un proceso continuo entre estudiantes y docentes que ayuda a comunicar los aciertos y las fallas. En ambos casos es una oportunidad para aprender y mejorar y no solo medir.

A fin de asegurar que la evaluación sea adecuada, es imprescindible preguntarse: ¿a quién se evaluará?, ¿en qué se le evaluará?, ¿cómo se evaluará? y ¿por qué razones se evaluará?

Con la finalidad de utilizar apropiadamente los instrumentos cuantitativos o cualitativos se requiere de un conocimiento de sus características, la forma en la que se diseñan y su contribución al proceso de enseñanza-aprendizaje.

En este marco, la rúbrica es un instrumento de evaluación que tiene cada vez más presencia en los espacios educativos, por la versatilidad de su uso y aplicación en múltiples

escenarios de aprendizaje. El propósito de este capítulo es poner al alcance del profesorado información que lo oriente en el diseño y el uso de este instrumento en la evaluación del y para el aprendizaje. El contenido está organizado en ocho secciones. La primera explica qué es una rúbrica, la segunda presenta los pasos para diseñarla, la tercera da indicaciones de cómo instrumentarla, la cuarta describe algunos aspectos de este instrumento en la evaluación a distancia, la quinta explica cómo analizar los resultados, la sexta incluye algunos ejemplos, la séptima presenta las conclusiones y reflexiones finales y la octava recomienda recursos en línea.

Este capítulo es un recurso de autoestudio, reflexión y consulta que puede leerse o no en orden. Pretende ser un punto de partida para que usted explore las posibilidades que ofrecen las rúbricas, practique su diseño, analice sus prácticas de evaluación e intercambie ideas con sus colegas.

¿QUÉ ES?

- La rúbrica es un instrumento que define tareas, actividades o comportamientos específicos que se desean valorar, así como los niveles de desempeño asociados a cada uno de estos. Es una guía articulada y precisa que ilustra los objetivos de cada tarea y su relevancia en el proceso de evaluación.
- Permite obtener una valoración objetiva basada en criterios determinados con anterioridad y conocidos por quienes intervienen en la evaluación, es decir, no pueden ser modificados caprichosamente.
- Existen dos tipos de rúbrica: *holística*, que proporciona un solo puntaje basado en la valoración integral del desempeño del alumnado y *analítica*, la cual arroja un puntaje diferente por cada componente de la rúbrica. La primera brinda una visión global del avance del alumnado, mediante resultados generales, mientras que la segunda permite valorar aspectos específicos del desempeño, a fin de identificar aciertos y errores.
- Usualmente se presenta en forma de matriz, con filas y columnas. En estas se incluyen los descriptores de lo que se desea evaluar –características, tareas o comportamientos– y niveles de desempeño. También puede incluir categorías que organizan un conjunto de descriptores afines.
- La escala de valoración depende del nivel de alcance determinado, la cual puede ser numérica; de frecuencia: nunca, casi nunca, algunas veces, frecuentemente, siempre; de adecuación: inadecuado, poco adecuado, adecuado y muy adecuado; de suficiencia: nulo, insuficiente, suficiente y de calidad de desempeño: muy bajo, bajo, medio, alto, muy alto/deficiente, regular, bueno, excelente/novata(o), aprendiz, experta(o). En general, poseen tres o cuatro niveles de desempeño.
- Una de sus ventajas, es que ayuda a contar con información precisa para expresar comentarios al alumnado sobre su rendimiento –con base en los avances logrados–, mientras que a la o el estudiante le permite identificar con claridad los elementos que requiere para cumplir con éxito las actividades descritas y también verificar sus tareas o trabajos

antes de entregarlos. En cuanto a sus limitaciones, la elaboración de la rúbrica requiere de tiempo, sobre todo, en dos etapas: primero, para planear y diseñar la rúbrica (escala de valoración, categorías o descriptores asociados a estas) y, después, para verificar la coherencia y concordancia entre estos elementos. Además, puede acotar el desempeño estudiantil por la descripción exhaustiva de los resultados esperados.

- En ambientes digitales se usa mucho porque ayuda a involucrar al alumnado en las tareas. Esto puede contribuir a disminuir el abandono por la guía que les ofrece en el proceso de aprendizaje.
- Puede ser empleada con propósitos de diagnóstico, por ejemplo, para verificar habilidades o conocimientos en el alumnado al inicio de un curso; con propósitos formativos, para valorar el avance en un tema en particular y diseñar actividades de apoyo, así como con propósitos sumativos, para explorar los conocimientos adquiridos al final de una unidad académica, a fin de juzgar el desempeño en un proyecto o tarea específica o con el objetivo de valorar el nivel de logro al concluir el ciclo escolar.
- Para quienes participan en la evaluación, la rúbrica puede ser un desafío que mantiene su atención y motivación.
- Es una guía que permite al personal docente calificar el desempeño de cada estudiante con los mismos criterios y simultáneamente obtener información sobre la efectividad de sus actividades de enseñanza.
- Si se utiliza adecuadamente, es una herramienta que brinda una visión detallada del proceso de aprendizaje del alumnado. En la siguiente sección se precisan los pasos para su diseño.

¿CÓMO LA DISEÑO?

La rúbrica es un instrumento muy útil en la valoración del aprendizaje, ya que puede implementarse en distintos escenarios y etapas de evaluación. Es importante dedicar tiempo a su diseño, tanto de contenidos como de estructura, para que pueda ser empleada adecuadamente y valorada de manera oportuna. A continuación, se aborda cómo desarrollarla.

Planeación

1) Seleccione el o los objetivos de aprendizaje que desea evaluar.
   ¿La rúbrica es el mejor instrumento para valorar ese aprendizaje, conducta o producto?
2) Considere el tipo de evaluación que quiere realizar con la rúbrica.
   Diagnóstica, formativa o sumativa.
3) Contemple a quién se dirigirá la rúbrica.
   Estudiantes, grupos de estudiantes, usted mismo o un grupo docente.
4) Determine el tipo de rúbrica.
   Analice el o los objetivos y con ello decida qué tipo de rúbrica elaborará.
   Holística.
   Ver ejemplo 1.

Analítica.
Ver ejemplo 2.

5) Defina la escala de valoración.
Elija una escala para valorar el desempeño del alumnado: frecuencia, adecuación, suficiencia, numérica o calidad de desempeño.
6) Establezca categorías generales.
Organice los descriptores de acuerdo con su afinidad y establezca categorías generales. Deben reflejar con claridad las dimensiones que cada estudiante demostrará en la evaluación.
7) Determine los descriptores para valorar el desempeño.
Con base en los objetivos defina las actividades o aspectos que incluirá en la rúbrica. Deben reflejar a detalle el nivel de logro que se espera del alumnado. Su redacción no debe dar lugar a interpretaciones.

Diseño

1) Defina el título de la rúbrica.
Corto, sencillo y atractivo.
2) Incluya una sección de datos de identificación.
Nombre de la asignatura, temática que aborda, datos de la o el docente, datos de la o el estudiante y fecha. Si emplea una plataforma para su diseño o la desarrolla dentro de *Google Classroom*, *Blackboard* o *Moodle*, algunos de estos elementos no serán necesarios.
3) Redacte las instrucciones para su uso.
Explique con claridad y precisión las condiciones para el empleo de la rúbrica (planeación, tiempo, evaluación).
4) Diseñe una tabla.
Holística.
Dos columnas
Una columna para la escala de valoración y otra para los aspectos específicos que desea evaluar.
Ver ejemplos 1 y 2.
Analítica.
Varias columnas.
Una columna con las categorías para valorar el desempeño y otras para la escala de valoración.
Ver ejemplos 3, 4 y 5.
5) Incluya la escala de valoración.
Holística.
Organícela de mayor a menor nivel, en una columna del lado izquierdo.
Analítica.
Asigne una columna para cada nivel. Coloque del lado derecho el menor y a la izquierda el mayor.

6) Incluya las categorías generales para valorar el desempeño.
   Organícelas de acuerdo con su importancia, de mayor a menor nivel, en una columna del lado izquierdo.
7) Añada los descriptores para valorar el desempeño
   Holística.
   Inclúyalos en una columna del lado derecho. Deben ubicarse de acuerdo con cada nivel en la escala de valoración.
   Analítica.
   Incorpore las tareas o aspectos que se desean evaluar de acuerdo con las categorías generales y con la escala de valoración definida.
8) Contemple un espacio para su valoración.
   Se puede añadir al final de la rúbrica para realizar una valoración global, o también, como columna adicional, para otorgar una valoración específica a cada categoría. Esto no es necesario si emplea una plataforma en línea.
9) Considere un apartado para observaciones.
   Inclúyalo al final de la rúbrica para que sus estudiantes expresen las dudas, inquietudes o comentarios que tengan sobre las actividades o aspectos contemplados. O para que usted exprese comentarios que ayuden a cada estudiante a comprender los aspectos que debe mejorar.
10) Edite la versión final.
   Revise la rúbrica y haga los ajustes para su edición final, ya sea que la comparta a través del correo electrónico o se despliegue en una plataforma o aplicación electrónica.

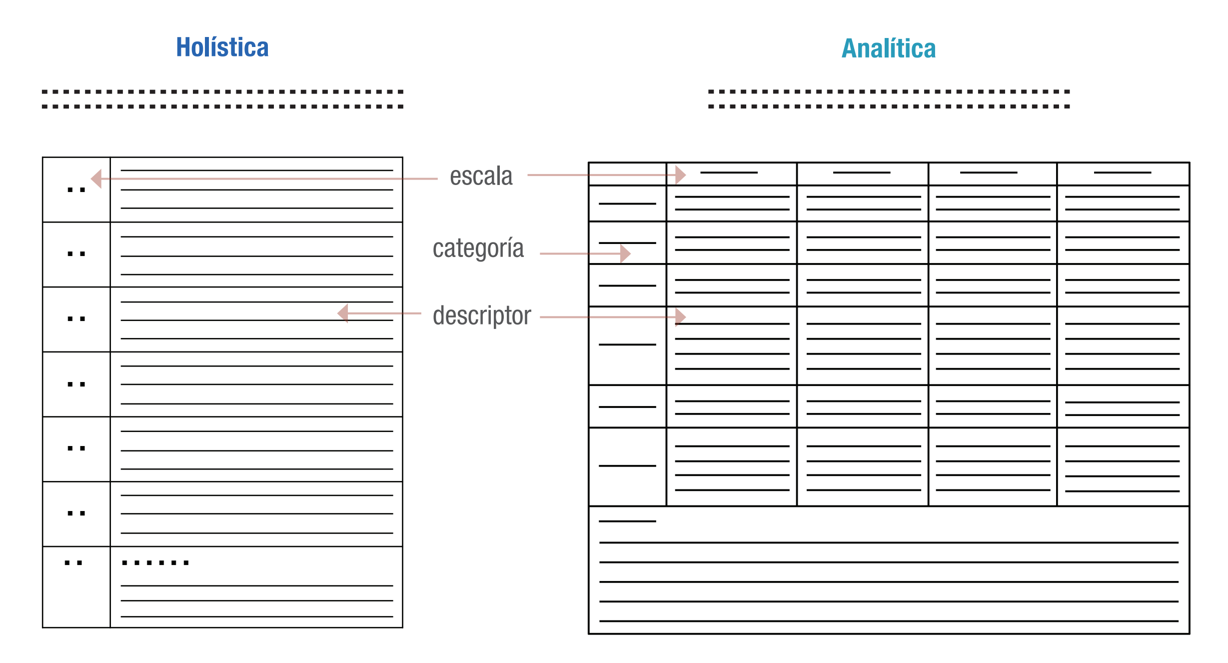
Consulte el video *Evaluar mediante rúbricas* https://www.youtube.com/watch?v=phpwpQuPFcA (Instituto de Docencia Universitaria, 2015).

Notas:
Defina los descriptores del nivel de desempeño más alto y después del más bajo, esto servirá de referencia para describir los niveles intermedios.

Revisión

- Es recomendable que revise periódicamente la rúbrica para asegurar su vigencia y pertinencia como instrumento de evaluación.
- Una vez aplicada es muy útil que pida a sus estudiantes comentarios o sugerencias sobre esta, dado que son quienes la utilizan, su opinión es muy importante para mejorarla.
- También puede compartirla con sus colegas, e inclusive especialistas para que valoren la pertinencia, claridad y organización de los aspectos a evaluar, en relación con los niveles de desempeño, además de su diseño.

Figura 1. Componentes de la rúbrica



Notas:

- En la rúbrica holística, el desempeño obtenido por cada estudiante puede ayudar a que identifique, de manera inmediata, los aspectos que puede mejorar en las actividades descritas.
- En la rúbrica analítica se pueden valorar los aspectos por categoría o por descriptor o de forma global (suma de categorías y descriptores).

¿CÓMO LA APLICO?

- De acuerdo con el tema abordado y el propósito de la evaluación, esta puede ser individual (una alumna o alumno o usted) o bien colectiva (un conjunto de estudiantes, todo el grupo o un equipo docente).
- Se puede incorporar fácilmente a las actividades de enseñanza y aprendizaje. Por ejemplo, estudiantes y docentes la pueden emplear para valorar el logro de un determinado desempeño o la calidad de un producto o de manera grupal para valorar exposiciones orales o trabajos en equipo.
- Con una rúbrica holística, la o el docente puede visualizar desde una perspectiva global el desempeño de sus estudiantes, mientras que, con una analítica, se obtiene un panorama más detallado de las razones por las que obtienen una determinada puntuación. En cualquiera de sus usos, permiten que docentes y estudiantes detecten de manera precisa los aspectos que requieren de revisión y mejora, casi de manera inmediata.

- Por la multifuncionalidad de la rúbrica, su aplicación puede adaptarse a distintos espacios educativos, tanto presenciales (como el salón de clases, laboratorio, taller), como virtuales (plataformas educativas, aplicaciones); la retroalimentación en estos últimos suele ser más rápida y sencilla, puesto que los sistemas disponen de múltiples opciones de respuesta, dependiendo del desempeño obtenido por cada estudiante.
- Las rúbricas ayudan mucho al alumnado en entornos virtuales de aprendizaje, dado que clarifican la realización de las actividades y contrarrestan la incertidumbre que genera la falta de interacción personal y de comunicación sincrónica que caracteriza estos espacios.
- La rúbrica se puede diseñar en versión digital para usarse en un dispositivo electrónico –teléfono inteligente, tableta o computadora personal– o para distribuirse en un archivo por correo electrónico, para lo cual se pueden emplear aplicaciones o páginas web diseñadas con este propósito (ver sección Recursos en línea). También se puede imprimir en papel, pero es importante incorporar un espacio específico para expresar observaciones o comentarios acerca de su contenido.
- No siempre la primera versión de una rúbrica es la mejor en todos sus aspectos, sin embargo, suele ser una herramienta que se mejora en la medida en que se utiliza y se detectan omisiones o ajustes necesarios en la definición de sus características. Para ello, es importante solicitar al alumnado una retroalimentación sobre la rúbrica, de manera que pueda ser revisada y afinada para asegurar su pertinencia.
- Si la rúbrica se desarrolla de forma adecuada y cumple con todas las características necesarias según su tipo, puede utilizarse con éxito en distintos ambientes de aprendizaje. Si el grupo tiene claridad sobre lo que debe hacer, no requerirá de la o el docente para su resolución. No obstante, debe haber una comunicación cercana para aclarar las dudas.
- Al aplicar la rúbrica, es necesario definir en qué momento será entregada, por ejemplo, se puede hacer antes de una actividad para que las y los estudiantes identifiquen el resultado de aprendizaje y tengan una guía para organizar su trabajo, o al final, para que valoren los niveles de desempeño y dimensiones en la ejecución de una actividad o en la elaboración de un producto. En ambos casos, permite que las y los estudiantes identifiquen sus aciertos o dificultades y, en función de ello, tomen decisiones para mejorar su aprendizaje.
- Aunado a esto, es importante determinar las condiciones para su empleo y devolución, que implica: si se resolverá de manera individual o colectiva (en equipos de trabajo o todo el grupo); los tiempos de entrega, ya sea al término de la actividad o después de un periodo determinado (corto, mediano plazo); así como los medios que se emplearán para compartirla (físicos o electrónicos). Se sugiere el uso de la rúbrica no solo como instrumento de evaluación, sino como una herramienta de reflexión para que las y los estudiantes tomen conciencia sobre lo que han aprendido.

Notas:

- Dedique tiempo a discutir la rúbrica con sus estudiantes antes de que hagan su trabajo, así podrán autoevaluarse con los criterios en los que colaboraron.
- Antes de que las y los estudiantes le entreguen sus trabajos finales pídales que los intercambien y hagan comentarios a sus pares usando la rúbrica.
- Use la rúbrica para que el alumnado conozca la expectativa que usted tiene sobre su desempeño.
- Pida a sus estudiantes que apliquen la rúbrica para evaluar productos de muestra antes de que elaboren los suyos.

LA RÚBRICA EN LA EVALUACIÓN A DISTANCIA

La modalidad de enseñanza a distancia demanda esfuerzos adicionales de planeación. Todas las actividades, incluida la evaluación, deben estar definidas con anticipación para dar mayor autonomía al alumnado. Por esa razón es necesario definir cómo se diseñará e instrumentará.

¿Cómo la diseño?

Una de las ventajas de emplear las rúbricas en la evaluación a distancia es que las herramientas en línea facilitan su desarrollo, difusión y calificación.

Ya sea que diseñe la rúbrica o que emplee alguna de las que están disponibles en algunas herramientas en línea, tiene la posibilidad de modificar los elementos que la integran. Es fácil distribuirla en alguna plataforma educativa o a través del correo electrónico. Además, en dichas plataformas, calificar la rúbrica es un proceso automatizado.

Ello ayuda no solo al profesorado, quien puede analizar los resultados individualmente, comparar el desempeño entre estudiantes o conocer el promedio del grupo casi de manera automática, sino también a cada estudiante para asumir la responsabilidad de su aprendizaje.

¿Cómo la instrumento?

La comunicación en la educación a distancia es vital, en este sentido, disponer de una rúbrica en la que se especifique con precisión la expectativa sobre el desempeño del alumnado hace posible dicha comunicación y ayuda a reducir su ansiedad, derivada del proceso evaluativo.

Es muy importante que cada docente defina si la evaluación será sincrónica o asincrónica, por ejemplo, si evalúa la presentación de un grupo de estudiantes lo recomendable es que los evalúe mientras exponen y al final de la sesión puedan discutir los resultados, pero si se trata de la entrega de un proyecto, lo más recomendable es la evaluación asincrónica, es decir, después de que se haya entregado el producto a evaluar. En ambos momentos puede involucrar al alumnado para que participe en la evaluación, pero es necesario que fije claramente las reglas de las intervenciones.

La rúbrica servirá de guía para que retroalimente al alumnado, puede agregar comentarios escritos o bien a través de un video o notas de voz, esos elementos sin duda animarán a las y los estudiantes.

Es importante que se asegure que todo el grupo comprenda la rúbrica, puede preguntar si la revisaron, si hay algún elemento que sea confuso, si la consideran útil para guiar su aprendizaje, entre otros aspectos.

¿Cómo diseño una rúbrica en *Rubistar*?

La herramienta *Rubistar* permite crear rúbricas de manera gratuita y rápida. Es posible guardarlas, editarlas e ingresar a ellas desde cualquier dispositivo con acceso a internet. A continuación, se presentan los pasos básicos para emplearla:

- Ingrese al siguiente enlace http://rubistar.4teachers.org/index.php
- Seleccione en el idioma de su preferencia (español o inglés).
- Podrá tomar la decisión de registrarse o no en la herramienta.
- Si decide registrarse, coloque sus datos (nombre, contraseña, correo electrónico) y oprima "Nueva rúbrica" para comenzar a utilizar la herramienta.
- Seleccione un tema de interés, de diferentes áreas, con base en las plantillas que se encuentran en la parte inferior de la página de inicio.
- Una vez que haya elegido un tema, se abrirá una nueva pantalla con un conjunto de diferentes rúbricas, seleccione una para comenzar a trabajar.
- Al escoger cualquier opción se desplegará un formato con la descripción de la rúbrica e información que es necesario añadir (título, código postal).
- Dispone de 40 minutos para que el sistema guarde la información de su rúbrica, si necesita más tiempo pulse el botón "Enviar y regresar a la página".
- En la parte inferior de la página se presenta la rúbrica en forma de matriz. Seleccione las categorías que requiera evaluar. Por ejemplo, si evalúa la comunicación oral, las categorías serán: contenido, volumen, vocabulario, comprensión, límite-tiempo. También puede cambiar el nombre de la categoría, así como los números por una escala cualitativa.
- Al finalizar presione la opción "Enviar" para exportar la rúbrica en distintos formatos (*Word* o *Excel*) o bien para imprimir la rúbrica.
- Si tiene dudas consulte la sección de preguntas frecuentes para identificar los aspectos esenciales para diseñar una rúbrica en esta herramienta.

Nota:

- Para conocer más detalles de esta herramienta consulte la sección de Recursos en línea al final de este capítulo.

Ejercicio de autoevaluación

Instrucciones: Marque con una X los elementos que cumplió al emplear *Rubistar* para diseñar su rúbrica, responda con sinceridad. Después sume los puntos que marcó e identifique la calificación que obtuvo en el ejercicio.

Tabla 1

☐	Exploró el área de trabajo y los distintos temas de las áreas de su interés.
☐	Definió el título de la rúbrica.
☐	Precisó la escala (cuantitativa o cualitativa) de acuerdo con las necesidades de evaluación.
☐	Detalló las categorías que mejor se ajustan a las necesidades de evaluación.
☐	Redactó con claridad y precisión el contenido de los descriptores.
☐	Se aseguró de que los descriptores fueran claramente diferenciables de acuerdo con el nivel de desempeño.
☐	Revisó las características generales de la lista (visualización, organización).
☐	Verificó la calidad del contenido (ortografía, presentación).
☐	Exportó o descargó la rúbrica.
☐	Compartió la rúbrica con sus estudiantes.
☐	Accedió al manual de usuario de la herramienta.

Puntuación	Desempeño
11	Excelente
10	Muy bueno
9	Bueno
8	Regular
7	Deficiente
6 o menos	Insuficiente

Fuente: Elaboración propia.

¿Cómo la analizo?

- Para analizar los resultados de las rúbricas es preciso considerar si se trata de una evaluación a cargo de la o el docente, una autoevaluación del alumnado o una coevaluación entre estudiantes; y también, el grado de participación: si se presentó antes de la actividad, para utilizarla como guía; después de la actividad, junto con la revisión y retroalimentación; o fue de uso exclusivo de la o el docente, sin presentarla a sus estudiantes.
- Cuando la usa la o el docente, facilita la revisión pues reduce la subjetividad y ambigüedad; al compartirla con sus estudiantes ayuda a que, desde el inicio, se definan criterios de evaluación y se use la rúbrica como guía de las actividades. Independientemente de su aplicación, la rúbrica permite que docentes y estudiantes tengan criterios o referentes comunes para la evaluación durante el desarrollo o ejecución de la tarea.
- Sin importar el tipo de rúbrica, usted tiene que considerar distintas posibilidades de respuesta para medir el nivel de desempeño del alumnado. De acuerdo con el tipo de evaluación y objetivos debe definir una escala para interpretar los resultados, para ello es necesario definir la información que requiere y cómo la utilizará una vez que la obtenga.
- Por ejemplo, en una rúbrica holística con una escala de valoración de cuatro niveles, usted puede determinar que el alumnado que cumpla con todas las características del cuarto nivel equivalga a un desempeño sobresaliente, mientras que quien efectúe completamente lo inherente al tercer nivel equivalga a un desempeño bueno y quien efectúe íntegramente lo concerniente a los niveles dos y uno corresponda a un desempeño regular e insuficiente, respectivamente. Para evaluar la rúbrica tendrá que definir una puntuación asociada a cada nivel de desempeño, lo que permitirá asignar una calificación al alumnado.
- En una rúbrica analítica también se puede emplear la misma escala de valoración, aunque en este caso se utilizará para evaluar aspectos más específicos, en cada uno se obtendrá un nivel de desempeño distinto. Por ejemplo, cuando se evalúa la participación en una exposición oral, que involucra un conjunto de categorías (dominio del tema, estructura, dicción…) y descriptores en particular. De esta forma, la o el docente obtiene una perspectiva particular y global del desempeño estudiantil.
- Usted puede aplicar una rúbrica holística al inicio del curso para evaluar las fortalezas y debilidades de cada estudiante y comparar su desempeño de acuerdo con el nivel esperado. En un segundo momento puede asignar equipos de trabajo de acuerdo con los resultados obtenidos. Una rúbrica analítica permitiría evaluar al grupo para visualizar lo que se puede mejorar y establecer un plan de acción basado en la información recabada.
- Los descriptores y niveles de desempeño de la rúbrica son muy útiles porque guían la valoración del desempeño del alumnado, de tal forma que cuando usted o el alumnado contrastan un producto, actividad o comportamiento con la rúbrica, obtienen un resultado inmediato. Además, puede documentar y medir el progreso de sus estudiantes.
- La rúbrica ofrece múltiples oportunidades de aprendizaje, ayuda al profesorado a identificar distintos aspectos del rendimiento de sus estudiantes, le permite valorar elementos

de la comunicación que favorecen u obstaculizan el aprendizaje y revisar la claridad en la definición de los criterios de evaluación; al alumnado, le permite contar con un referente claro en caso de que tenga dudas sobre su desempeño.

- En la retroalimentación es recomendable que, además de indicar el nivel de desempeño obtenido, se le proporcione al alumnado una breve explicación que atienda a los puntos que es necesario mejorar en su desempeño. Comentarios generales si se trata de una rúbrica holística o síntesis de los elementos más importantes, si se trata de una analítica.

Ejemplos

1) Debate

Instrucciones: Elija el nivel de desempeño de cada estudiante en el debate y asigne la calificación correspondiente.

Tabla 2

5	Los argumentos que expone son excelentes y los expresa con gran claridad y precisión, muestra mucha seguridad en sus intervenciones y se conduce en todo momento con respeto ante la audiencia.
4	Los argumentos que presenta son muy buenos y los expresa con claridad y precisión, demuestra seguridad en sus intervenciones y respeta a la audiencia.
3	Su argumentación es buena y la expone claramente, pero falta precisión; demuestra seguridad al intervenir y trata con respeto a la audiencia.
2	Los argumentos que emplea son poco sólidos y falta más precisión y claridad al comunicarlos, en varios momentos muestra inseguridad y trata con respeto a la audiencia.
1	Los argumentos que emplea son deficientes y los expresa de forma confusa e imprecisa, demuestra poca seguridad al intervenir y muestra poco respeto por la audiencia.
0	Los argumentos son incorrectos y los expone de manera confusa e imprecisa, demuestra mucho nerviosismo en sus intervenciones y no respeta a la audiencia.
Valoración	**Observaciones**

Puntuación	Desempeño	Calificación
5	Excelente	10
4	Muy bueno	9
3	Bueno	8
2	Regular	7
1	Deficiente	6
0	Insuficiente	5

Fuente: Elaboración propia.

2) Comunicación escrita

Instrucciones: Elija en cada categoría el nivel de desempeño de la alumna o el alumno y anote su puntuación en la columna de la derecha. Cada categoría tiene un peso diferente; para calcular el porcentaje final identifique en la tabla de abajo el porcentaje que le corresponde a la puntuación de cada categoría y anótelo en la última columna. Sume el porcentaje de todas las categorías para determinar la calificación de la alumna o el alumno de acuerdo con la escala ubicada al final de la rúbrica.

Tabla 3

Categoría/ Desempeño	Deficiente 1	Regular 2	Bueno 3	Excelente 4	Puntuación	%
Tratamiento del tema 50%	No aborda los conceptos principales ni expone sus características.	Menciona solo algunos conceptos y expone algunas de sus características.	Menciona la mayoría de los conceptos centrales y expone sus características.	Profundiza en los conceptos centrales y expone sus principales características.		
Organización 20%	El contenido está desorganizado y es muy difícil seguir la secuencia de ideas.	El texto tiene fallas en la secuencia lógica que impiden comprenderlo completamente.	La secuencia del contenido es lógica, pero tiene algunas ideas incompletas.	El contenido se presenta en una secuencia lógica que facilita su comprensión.		
Redacción 20%	El texto presenta muchas fallas en su redacción.	El texto contiene fallas de redacción.	La mayor parte del texto está escrito correctamente.	El texto está escrito correctamente.		
Ortografía 10%	Contiene demasiados errores de ortografía y puntuación.	El texto contiene algunos errores ortográficos y de puntuación.	El texto casi no tiene errores ortográficos y de puntuación.	No contiene errores de ortografía y puntuación.		
Observaciones					**Porcentaje total**	

	Tratamiento del tema	Organización	Redacción	Ortografía
4	50%	20%	20%	10%
3	37.5%	15%	15%	7.5%
2	25%	10%	10%	5%
1	12.5%	5%	5%	2.5%

Porcentaje	Desempeño	Calificación
100%	Excelente	10
90-99%	Muy bueno	9
80-89%	Bueno	8
70-79%	Regular	7
60-69%	Deficiente	6
50-59%	Insuficiente	5

Fuente: Elaboración propia.

CONCLUSIONES Y REFLEXIONES FINALES

- La rúbrica es un instrumento que define tareas, actividades o comportamientos específicos que se desean valorar, así como los niveles de desempeño asociados a cada uno de estos, la cual permite obtener una valoración objetiva y precisa sobre el rendimiento del alumnado.
- Existen dos tipos de rúbrica: la holística y la analítica. La primera brinda una perspectiva global del desempeño del estudiantado y la segunda ofrece evidencia más detallada y específica sobre cada aspecto evaluado. El empleo de una u otra depende de los objetivos de la evaluación.
- Usualmente se presenta en forma de matriz, con los aspectos que se desean evaluar (descriptores) y los niveles de desempeño asociados a cada uno. Además, se pueden emplear múltiples escalas de valoración, aunque es importante que cada rúbrica se diseñe con una sola escala.
- Puede ser empleada en diferentes momentos del aprendizaje y dirigida no solo al alumnado, sino también a una profesora o profesor o a un equipo docente. Puede diseñarse de forma que permita comparar el desempeño.
- Dependiendo de las características del curso, se puede diseñar en versión digital para utilizarse en dispositivos electrónicos o imprimir en papel, para compartirla directamente con el alumnado.
- Es posible emplearla en múltiples espacios de aprendizaje, tanto presenciales como virtuales, lo que ayuda a involucrar al alumnado en las tareas y a asumir su responsabilidad en el proceso de aprendizaje.
- Aunque su diseño requiere tiempo y puede resultar complejo la primera vez que se realiza, facilita la recuperación rápida de resultados de cada estudiante lo que permite que el profesorado valore el desempeño individual, pero también, juzgue los resultados de todo el grupo.
- Al emplearla, se dispone de referentes comunes de evaluación, lo que favorece el aprendizaje autónomo y la autorregulación del alumnado, así como una mejor planeación del curso para el profesorado.
- Es deseable que antes de diseñar una rúbrica determine si es el instrumento más idóneo para medir el aprendizaje de sus estudiantes y si proporcionará evidencias útiles.
- Se sugiere revisar periódicamente la rúbrica para verificar su vigencia y pertinencia con la ayuda de las y los destinatarios.

RECURSOS EN LÍNEA

Tabla 4

Recurso	Descripción
ForAllRubrics http://www.forallrubrics.com	Permite añadir el número de filas o columnas que se desee y asignar diferentes colores y puntajes a los niveles de desempeño. Dispone de una librería para explorar rúbricas elaboradas por otras personas que se pueden copiar o imprimir.
Rcampus http://www.rcampus.com/	Es posible diseñar cada elemento de la rúbrica o emplear un formato existente. Se pueden modificar sus diferentes componentes y evaluarlos en poco tiempo. Almacena las evaluaciones en el área de trabajo y permite compartirlas.
Quick Rubric https://www.quickrubric.com/	Permite añadir, editar y eliminar filas o columnas y organizarlas de acuerdo con diferentes criterios. Dispone de un espacio para la descripción de la rúbrica y sus instrucciones. Asigna puntajes máximos y mínimos de calificación, cuyos cambios ajustan automáticamente el valor de las categorías.
Evalcomix http://evalcomix.uca.es/	Dispone de una interfaz que permite modificar todos los elementos de la rúbrica: título, dimensiones y aspecto, valoración global y específica, espacio de comentarios. Incluye un manual detallado para la elaboración de la rúbrica.
CoRubric http://corubric.com/	Ayuda a crear rúbricas que pueden ser compartidas de manera privada o pública. Permite asignar ponderaciones de forma manual o a partir de criterios definidos por cada persona. Es posible exportar las rúbricas en diferentes formatos.
Erubrica http://www.erubrica.com/	Ayuda a crear rúbricas o elegirlas de un banco, ajustándolas de acuerdo con los aspectos que se desea evaluar. Se pueden resolver directamente en la página y compartirlas. Además, permite almacenar las evaluaciones realizadas y descargar las rúbricas en diferentes formatos.
Rubistar http://rubistar.4teachers.org/index.php	Ayuda a crear rúbricas en minutos, añadir filas y columnas, ajustar su contenido y acceder a una vista previa de su presentación. Se pueden buscar rúbricas elaboradas por otras personas entre una gran diversidad de temas e imprimirlas para uso personal. Cuenta con un manual de ayuda.
The rubric builder http://www.rubricbuilder.com/	Permite diseñar una rúbrica o elegir de un catálogo, editar su contenido y generar los niveles de desempeño de manera automática. Es posible imprimir o exportar en diferentes formatos.
The rubric maker https://rubric-maker.com/	Permite asignar a las rúbricas título, subtítulo y descripción, así como un puntaje para cada tema. Incorpora cuatro columnas con diferentes niveles de desempeño que se ajustan de acuerdo con los descriptores determinados por cada persona.

REFERENCIAS

Allen, J. (2014). *Using Rubrics to Grade, Assess, and Improve Student Learning.* Miami, Estados Unidos: Miami Dade College. Recuperado de https://www.academia.edu/39389988/Using_Rubrics_to_Grade_Assess_and_Improve_Student_Learning

Andrade Restrepo, L. M., Andrade Restrepo, M. A. (2017). Uso de las rúbricas en ambientes virtuales de aprendizaje. *Revista Educación y Humanismo*, *19*(32), pp. 102-110. Recuperado de http://revistas.unisimon.edu.co/index.php/educacion/article/view/2535

Barrios Ríos, O. (2018). *Uso de rúbricas. Una guía para el profesor*. Universidad Tecnológica Metropolitana. Vicerrectoría Académica. Unidad de Mejoramiento Docente. Recuperado de https://vrac.utem.cl/wp-content/uploads/2018/10/Manual.Uso_Rubricas.-2.pdf.

Bijl, R. (s.f.). *Manual Rubrics.* Recuperado de: http://www.hum2.leidenuniv.nl/ECOLe/story_content/external_files/Handleiding%20Rubrics.pdf

Blackboard 9.1. (2013). *About Rubrics and why to use them.* Queensland, Australia: University of the Sunshine Coast (USC). Center for support and advancement of learning and teaching. Recuperado de https://www.usc.edu.au/media/2737688/About-Rubrics.pdf

Blackboard Learn. (2012). *Using Rubrics.* Recuperado de https://help.online.uts.edu.au/wp-content/uploads/2015/05/using_rubrics.pdf?e7ad49

Brophy, T. S. (s.f.). *Writing Effective Rubrics.* Estados Unidos: University of Florida Institutional Assessment. Recuperado de http://web.cse.ohio-state.edu/~soundarajan.1/abet/writing_effective_rubrics_guide_v2.pdf

Cano, E. (2015). Las rúbricas como instrumento de evaluación de competencias en educación superior: ¿Uso o abuso? Profesorado. *Revista de Currículum y Formación de Profesorado,* 19, 265–280. Recuperado de https://www.redalyc.org/pdf/567/56741181017.pdf.

Centurión, I., Lara M. y Román C. (2021). El uso de la rúbrica para evaluar tareas en entornos virtuales de aprendizaje: Un caso aplicado a estudiantes avanzados de la carrera de economía. *Revista d'Innovació Docent Universitaria, 13*, 9-17. Recuperado de https://revistes.ub.edu/index.php/RIDU/article/view/RIDU2021.13.2/33081

Dirección General Académica y Dirección de Currículo y Formación. (2020), Instrumentos de Evaluación ¿Cómo hacer una rúbrica? Universidad EIA. Recuperado de https://www.eia.edu.co/wp-content/uploads/2020/10/Como-hacer-una-rubrica.pdf

Gatica-Lara, F. y Uribarren-Berrueta, T. del N. J. (2013). ¿Cómo elaborar una rúbrica? *Investigación Educativa Médica*, *2*(1), 61-65. Recuperado de http://www.riem.facmed.unam.mx/index.php/riem/article/view/452

Instituto de Docencia Universitaria. (2015, 28 de abril). *Evaluar mediante rúbricas* [Video]. YouTube. https://www.youtube.com/watch?v=phpwpQuPFcA

Jonsson, A. y Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, *2*, 130–144. Recuperado de https://www.sciencedirect.com/science/article/abs/pii/S1747938X07000188

Kweksilber, C., Trías, D. (2020). Rúbrica de evaluación. Usos y aprendizajes en un grupo de docentes universitarios. *Páginas de Educación, 13*(2), pp. 100-121. Recuperado de http://www.scielo.edu.uy/scielo.php?script=sci_arttext&pid=S1688-74682020000200100.

Lara-Inostroza, F. y Cabrera-Pommiez, M. (2015). Ficha N° 20 Rúbricas. F. Lara-Inostroza y M. Cabrera-Pommiez, *Fichas de procedimientos de evaluación educativa UDLA*. Santiago de Chile, Chile: Universidad de las Américas. Recuperado de http://www. udla.cl/portales/tp9e00af339c16/uploadImg/File/PlanesDeEstudio/Fichas-de-proce-dimientos-de-evaluacion-UDLA-b.pdf

Mertler, C. (2001). Designing Scoring Rubrics for Your Classroom. *Practical Assessment, Research & Evaluation, 7*(25). Recuperado de https://scholarworks.umass.edu/pare/vol7/iss1/25/

Miller, A. (2020). Formative Assessment in Distance Learning. Recuperado de https://www.edutopia.org/article/formative-assessment-distance-learning

Sánchez, M. M., y Martínez, G. A. (Eds.). (2020). Evaluación del y para el aprendizaje: instrumentos y estrategias (1a ed.). Ciudad de México, México: UNAM. Recuperado de https://cuaieed.unam.mx/descargas/investigacion/Evaluacion_del_y_para_el_aprendizaje.pdf

Zúniga, L. (2017). Las rúbricas o matrices de valoración. *Diseño y uso de rúbricas para evaluar competencias comunicativas. Una propuesta desde la evaluación formativa* (70-90). Cusco, Perú: Danny's Graff E.I.R.L. Cusco – Perú. Recuperado de https://www.academia.edu/39603353/Libro_final

Capítulo 16

ANÁLISIS PSICOMÉTRICO DE EXÁMENES TEORÍA DE MEDICIÓN CLÁSICA

Enrique Ricardo Buzo Casanova, Manuel García Minjares

"Hacer como el carpintero: medir dos veces, para cortar una vez."
(Refrán popular)

INTRODUCCIÓN

Sin duda, una de las consideraciones latentes al momento de diseñar un instrumento de evaluación, es que este arroje suficiente evidencia de validez que permita sustentar la interpretación de los resultados. Una pieza clave para alcanzar este objetivo es el análisis del instrumento, donde se puede constatar, por ejemplo: su dificultad, la factibilidad de los distractores, la calidad del diseño o la idoneidad de los niveles de respuesta, aspectos de diseño que son puestos a prueba cuando se someten a la inspección de este lente. Con el propósito de brindar criterios objetivos de evaluación, desde comienzos del siglo XIX, se buscó desarrollar métodos que permitieran de forma objetiva analizar la calidad psicométrica de diferentes instrumentos de evaluación, lo que se consolida a comienzos del siglo XX con el desarrollo de la *Teoría Clásica de los Tests* a partir de los trabajos de Spearman (Sartes, 2013) cuya aplicación, hoy en día, se extiende al análisis de exámenes objetivo. La Teoría Clásica de los Tests (TCT) es un análisis psicométrico que permite conocer el comportamiento estadístico de los reactivos que componen un examen como del instrumento en sí. Este tipo de análisis contribuye a que los docentes elaboren mejores instrumentos de evaluación que reflejen con mayor certeza el aprendizaje de sus estudiantes, además de que su aplicación es viable en todos los niveles educativos y de evaluación psicológica y hasta cierto punto su interpretación no resulta complicada.

La utilización emergente de la educación a distancia derivada del confinamiento por el COVID-19 mostró la necesidad de contar con métodos de enseñanza y evaluación que permitan realizar inferencias con evidencia de validez y el modelo de la TCT es una alternativa que permite alcanzar este objetivo.

En este capítulo se brindará las bases para analizar el funcionamiento de un examen objetivo con empleo de la TCT, y se habla sobre el Sistema de Análisis Psicométrico de Reactivos

(SISAPRE) diseñado en la Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia (CUAIEED) de la Universidad Nacional Autónoma de México para calibrar un examen conforme a la TCT.

¿QUÉ ES LA TCT?

La TCT es un conjunto de conceptos y técnicas que dan sustento al desarrollo de numerosos instrumentos de evaluación (DeVellis, 2006), se fundamenta en el modelo lineal propuesto por el psicólogo británico Charles Spearman (1863–1945) en el que se plantea que la puntuación empírica de un test (X) es resultado de un error de medición (e) sobre la puntuación verdadera (V) es decir

$$X = V + e \qquad (1)$$

Donde *V* y *e* son desconocidos.

Para fortalecer al modelo (1) se establecen los siguientes supuestos:

- El valor esperado de la puntuación empírica es la verdadera (E(X) = V), es decir, si a un estudiante se le aplicara muchas veces el mismo instrumento el valor promedio sería su puntuación verdadera, lo que implica que a la larga se espera que el error de medición desaparezca.
- No existe asociación entre la puntuación verdadera y el error de medición ($\rho_{Ve} = 0$), esto significa que las puntuaciones elevadas o bajas de un estudiante no son consecuencia de un error de medición.
- Los errores de medición en exámenes distintos no guardan relación($\rho_{e_j e_k} = 0$). Con esto no es de esperarse que estudiantes con errores de medición elevados en un examen lo vuelvan a presentar en otro.

El modelo también plantea como definición adicional el paralelismo de exámenes, es decir, la evaluación de un mismo constructo a través de instrumentos diferentes, lo que deriva en concebir a los exámenes paralelos como aquellos que miden lo mismo con diferentes reactivos. Por tanto, si el examen A es paralelo al B, sus puntuaciones verdaderas serán iguales ($V_A = V_B$) al igual que la variabilidad de sus errores de medición $(O_{e_A}^2 = \sigma_{e_B}^2)$.

A partir de este modelo clásico, en conjunto con sus supuestos, se deriva un resultado que es de mucha utilidad al momento de evaluar un instrumento: la confiabilidad.

Confiabilidad de un examen objetivo

Del modelo (1) puede observarse que conforme el error de medición es menor, la puntuación empírica (*X*) se acerca más a la verdadera (*V*), lo que implica que la confiabilidad de un examen radica en la cuantía del error de medición (*e*), sin embargo, existe la limitante de que

tanto la puntuación verdadera (V) como el error de medición (e) no se conocen. Para superar esta problemática se recurre a estimar de forma indirecta el error de medición (e) a partir de la correlación de dos exámenes paralelos. De acuerdo a la definición de formas paralelas, si ambos instrumentos se aplicaran a muestras de alumnos parecidas, se esperaría registrar las mismas puntuaciones verdaderas, lo que significaría tener una correlación perfecta, sin embargo, cualquier distanciamiento de esta situación, se debe al error de medición.

Supóngase que X y X' representan dos exámenes paralelos, la confiabilidad de X será la correlación de los exámenes paralelos ($\rho_{XX'}$) la cual es la proporción que representa la varianza de las puntuaciones verdaderas de las empíricas (Muñíz,1996), esto es:

$$\rho_{XX'} = \frac{\sigma_V^2}{\sigma_X^2} \tag{2}$$

de los supuestos del modelo (1), se desprende que:

$$\sigma_V^2 = \sigma_X^2 - \sigma_e^2 \tag{3}$$

al combinar (1) y (2) se obtiene:

$$\rho_{XX'} = \frac{\sigma_V^2}{\sigma_X^2} = \frac{\sigma_X^2 - \sigma_e^2}{\sigma_X^2} = 1 - \frac{\sigma_e^2}{o_X^2} \tag{4}$$

La expresión (4) muestra de manera más explícita que el nivel de confiabilidad de un examen está en función del error de medición expresado en su varianza, por otro lado, este resultado sirve de base para estimar el error típico de medición (SEM por sus siglas en inglés) y la puntuación verdadera (V).

Determinación del error típico de medición y la puntuación verdadera

Del modelo (1) se desprende que el error de medición es la diferencia entre la puntuación empírica y la verdadera, es decir, $e = X\text{-}V$, como ya se mencionó con anterioridad, V se desconoce, por lo que no es factible conocer e. En vez de intentar estimar el error de medición e se trabaja con la desviación estándar de esta variable (σ_e) a la cual se le conoce como el error típico de medición o SEM por sus siglas en inglés. Este error típico se estima al despejar (4) con la fórmula:

$$\sigma_e = \sigma_x\sqrt{1 - \rho_{XX'}} \tag{5}$$

donde la estimación depende de las puntuaciones empíricas.

Un supuesto que con frecuencia se incorpora al modelo (1) es que los errores de medición tienen una distribución normal con media 0 y varianza σ_e^2, de cumplirse lo anterior, entonces para algún valor *V*, la puntuación empírica sigue una distribución normal con media *V* y varianza σ_e^2, por tanto, la puntuación verdadera está contenida en el intervalo:

$$X \pm z_{1-\frac{\alpha}{2}}\sigma_e \qquad (6)$$

Donde $z_{1-\frac{\alpha}{2}}$ es el cuantil de una distribución normal estandarizada que garantiza un intervalo que contiene la puntuación verdadera con una confiabilidad de *1-α (0 < α < 1).*

Cálculo del coeficiente de confiabilidad

En el cálculo del error típico de medición que se determinó en (5) interviene el coeficiente de confiabilidad, el cual hasta el momento se considera como la correlación de dos exámenes paralelos, ahora bien, el cálculo de la confiabilidad de forma empírica puede llevarse a cabo con la aplicación del examen a la misma población en dos momentos diferentes (test–re test), o aplicar una mitad del examen a la mitad del grupo y la otra al resto (dos mitades) o utilizar formas equivalentes. Llevar a cabo alguna de estas dinámicas puede ocasionar inconvenientes de logística. Una alternativa para el cálculo de la confiabilidad de un examen objetivo con los resultados de la aplicación es la fórmula propuesta por Cronbach (1951), la cual se le denomina *alfa de Cronbach* cuyo valor normalmente es menor o igual al coeficiente de confiabilidad que se obtiene con exámenes equivalentes. (Muñíz, 1996). El *alfa de Cronbach* se calcula así:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\Sigma_{i=1}^{n} \ \sigma_i^2}{\sigma\frac{2}{x}}\right) \qquad (7)$$

donde n es el número de reactivos en el examen, σ_x^2 es la varianza de las puntuaciones empíricas σ_i^2 y es la varianza del i-ésimo reactivo. El valor de α dependerá de la manera en que se encuentren interrelacionados los reactivo s, obsérvese que la fórmula es semejante a (4). En la práctica se considera que un instrumento tiene una confiabilidad aceptable a partir de 0.7.

Análisis de los reactivos del examen

La TCT se enfoca tanto en el comportamiento del reactivo por sí mismo y su comportamiento como parte del instrumento en su totalidad. El análisis psicométrico de reactivos, llamado también calibración, engloba una serie de procedimientos con los que se evalúa, además del comportamiento del examen en sí, el funcionamiento de los reactivos y sus opciones de respuesta, lo que permite verificar de manera cuantitativa la dificultad del reactivo, su utilidad para determinar el grado de conocimiento de quienes contestan, además de inferir si su construcción favorece o complica su resolución. De esta manera, se cuenta con mayores elementos para el momento de tomar decisiones sobre el reactivo, si se tiene que cambiar,

mejorar o si es necesario eliminarlo del examen. En el análisis de los reactivos se revisan primordialmente tres aspectos: la dificultad, la discriminación y la correlación punto biserial.

Dificultad

La dificultad de un reactivo se denotará como *p* y se define como la proporción de alumnos que responden correctamente, se puede expresar así:

$$\rho_i = \frac{A}{N} \tag{8}$$

Donde:
p_i = Dificultad del i-ésimo reactivo (i=1,…,n).
A = Examinados que respondieron correctamente el reactivo. (A≤N)
N = Población expuesta al reactivo.

La dificultad de un reactivo puede tomar valores entre 0 y 1, mientras más cercano sea a uno significa que más alumnos lo respondieron de forma correcta y, por lo tanto, el reactivo es más fácil. Por el contrario, a medida que la dificultad se acerca a cero implica que una cantidad menor de estudiantes lo respondió correctamente, por lo que el reactivo es más difícil. Los reactivos pueden clasificarse de acuerdo a su dificultad, un criterio de agrupación es con una distribución uniforme, si la dificultad es menor a 0.2, el reactivo es muy difícil; entre 0.2 y 0.4 difícil; entre 0.4 y 0.6 regular; entre 0.6 y 0.8 difícil y mas de 0.8 muy fácil. En función del contexto de cada aplicación, es recomendable que un reactivo tenga un valor de dificultad que refleje la respuesta acertada de por lo menos la quinta parte de la población total, por ejemplo, un reactivo difícil debe tener al menos un 20% de respuestas correctas y, por el contrario, un reactivo fácil no debe exceder el 80% de dificultad.

Discriminación

La discriminación de un reactivo se traduce como su capacidad para distinguir entre los alumnos que cuentan con el conocimiento o habilidad respecto a los que no. Regularmente toma como base el 27% de los puntajes menores, a los que se les denomina el grupo bajo y el 27% de los mayores a quienes consideraremos como el grupo alto. Estos grupos se contrastan entre sí conforme a la siguiente fórmula:

$$D = \frac{A}{n_A} - \frac{B}{n_B} \tag{9}$$

Donde:
D=Discriminación.
A=Número de examinados del grupo alto que contestan bien al reactivo.
n_A=Número de examinados en el grupo alto.
B=Número de examinados del grupo bajo que contestan bien al reactivo.
n_B=Número de examinados en el grupo bajo.

La discriminación de un reactivo es la diferencia de la dificultad que observa en cada grupo, es decir, compara qué proporción de alumnos contestó correctamente el reactivo en cada grupo. Esta medida toma valores entre -1 y 1, es de esperarse que un reactivo bien elaborado sea contestado en mayor medida por los alumnos de mayor habilidad o conocimiento por lo que esta medida debiera ser, en un primer momento, positiva. Si la discriminación de un reactivo fuera negativa, es indicativo de que el funcionamiento del reactivo no es el adecuado porque lo contesta mejor el grupo de menor desempeño, y si tuviera un valor cercano a cero significa que el reactivo no está discriminando porque se contesta igual independientemente del nivel de habilidad de los alumnos. A partir de 0.2 se considera que el nivel de discriminación es aceptable.

Correlación punto biserial

Los coeficientes de correlación son medidas estadísticas que nos describen la relación lineal entre variables, toman valores entre -1 y 1. Un valor de correlación conforme más se acerque a 1 indica que los cambios en una variable afectan directamente a otra, conforme estos valores se acerquen a -1 indican lo contrario, cuando es cercano a cero indica que no existe relación lineal entre las variables.

Para el análisis de reactivos empleamos una correlación entre el puntaje del examen y la selección de la opción correcta del reactivo de interés. La fórmula resultante es la siguiente:

$$\rho_{pb} = \frac{\mu_A - \mu_x}{\sigma_x}\sqrt{\frac{p}{q}} \tag{10}$$

donde:

μ_A = puntaje promedio en el examen de los alumnos que contestaron correctamente el reactivo.
μ_x = puntaje promedio en el examen de todos los alumnos.
σ_x = desviación estándar del puntaje del examen.
p = dificultad del reactivo.
q = 1-p.

Conforme a la expresión (9), la correlación punto biserial es una distancia estandarizada entre las medias de aciertos de los alumnos que contestan correctamente el reactivo y la media de la población, ajustada por la dificultad del reactivo. Si la correlación punto biserial arroja un valor inferior a cero significa que el desempeño de los alumnos que contestan de forma correcta el reactivo es inferior al de la población global, lo que implica que el reactivo no está funcionando porque de alguna manera *se requiere no saber para contestarlo*. Si la correlación del reactivo es cero, significa que no existe diferencia en el desempeño de los alumnos que lo responden bien respecto al resto, o bien, el ítem es muy difícil. Finalmente, si la correlación punto biserial es positiva, es señal que los alumnos que eligieron la respuesta correcta tuvieron un mejor desempeño en comparación al general. Como se espera

que un buen reactivo contribuya de manera importante en el puntaje del examen el valor de la correlación punto biserial debe ser positivo. Con base a la evidencia empírica de la CUAIEED, reactivos con correlaciones a partir de 0.15 muestran un funcionamiento aceptable, es deseable tener valores de 0.2 en adelante. La correlación punto biserial en cierta forma evalúa la calidad del diseño del reactivo, la experiencia de la Coordinación ha demostrado que reactivos con correlaciones minúsculas o negativas, incluso, están asociados a errores en el diseño de la base del reactivo o de alguno de los distractores.

Análisis de los distractores

Para analizar el comportamiento de los distractores es suficiente calcular para cada uno su discriminación y correlación punto biserial. A diferencia de la opción correcta, se espera que los distractores sean elegidos por los alumnos de menor habilidad o conocimiento, por lo que la proporción de alumnos del grupo bajo que los eligen supera a los del alto. Por otro lado, también se espera que los alumnos que seleccionan los distractores tengan un puntaje menor que el de la población. Por tanto, los distractores en un reactivo que funciona correctamente tienen una discriminación y correlación punto biserial negativa.

Si se graficara el comportamiento de las opciones de respuesta de acuerdo a la proporción de alumnos de los grupos de desempeño que las eligen, se esperaría que la respuesta correcta se eligiera más a medida que mejora el desempeño de los examinados, mientras que los distractores se esperaría un comportamiento inverso. La Figura 1 muestra un ejemplo de un reactivo que funciona de manera deseable.

Figura 1. Ejemplo de comportamiento deseable de las opciones de respuesta de un reactivo

Calibración realizada en el Sistema de Análisis Psicométrico de Reactivos (SISAPRE).
Fuente: Dirección de Evaluación Educativa de la CUAIEED.

En el reactivo ilustrado en la Figura 1, la opción correcta es la C, el cual se selecciona más conforme la habilidad del examinado es mayor, por el contrario, los distractores son menos factibles de ser seleccionados conforme mejora el desempeño de los estudiantes.

Para afianzar los temas que se han comentado a lo largo de este capítulo, se muestra a continuación tres ejemplos de resultados de una calibración de acuerdo a la TCT.

Ejemplo 1. Análisis global del instrumento

La Tabla 1 presenta resultados de la calibración realizada al examen diagnóstico de conocimientos al ingreso al bachillerato de la UNAM de la generación 2020 (Sánchez Mendiola, M. et. al., 2020) conforme a la Teoría Clásica.

Tabla 1. Estadísticos del examen diagnóstico de ingreso al bachillerato Generación 2020

Estadístico	Valor
Coeficiente de confiabilidad (alfa)	0.825
Error típico de medición (SEM)	4.966
Media de dificultad	0.551
Media del coeficiente de correlación punto biserial	0.174

Fuente: Sánchez Mendiola, M. et. al. (2020). *Evaluación diagnóstica de conocimientos. Resultados de los alumnos que ingresan al bachillerato de la UNAM. 2020.* CODEIC, UNAM. P 13.

La Tabla 1, muestra en primer lugar al coeficiente de confiabilidad del examen, a través del *alfa de Cronbach*, cuyo valor es 0.825, lo que habla de que el instrumento es confiable por ser mayor a 0.7; el error típico de medición de 4.966 dice en general que los errores se encuentran a esa distancia de su valor esperado; la media de dificultad y del coeficiente de correlación punto biserial respectivamente expresan que el examen tuvo un nivel de dificultad intermedio, ya que se espera que 55% de los alumnos respondan correctamente algún reactivo del que se conforma el instrumento, con un funcionamiento aceptable además.

Ejemplo 2. Análisis de reactivos

La Tabla 2 presenta los parámetros de dos reactivos que se calibraron conforme a la TCT.

Tabla 2. Parámetros de dos reactivos conforme a la TCT

Parámetro	Reactivo 1	Reactivo 2
Dificultad	0.263	0.558
Discriminación	0.032	0.286
CPB	-0.010	0.197

CPB=Correlación punto biserial.

Fuente: Dirección de Evaluación Educativa de la CUAIEED.

La Tabla 2 exhibe a dos reactivos con comportamientos diferentes, el primero es un reactivo difícil porque lo contestó acertadamente el 26% de los alumnos; el valor de discriminación indica que no hay diferencia importante en la forma en que lo responden los alumnos del grupo alto respecto a los del bajo; la correlación punto biserial muestra que no existe una diferenciación del desempeño de los alumnos que contestan bien el reactivo respecto al global. En contraste, el segundo reactivo tiene un nivel de dificultad intermedio, discrimina de manera clara y los alumnos que eligen la opción correcta tienen un mejor desempeño en el examen respecto a los que se inclinan por un distractor.

Ejemplo 3. Análisis de los distractores

Supóngase que el reactivo 1 del ejemplo anterior cuenta con cuatro opciones de respuesta y la proporción de alumnos que selecciona cada una, así como su correlación punto biserial son las que muestra la Tabla 3.

Tabla 3. Proporción de alumnos y correlación punto biserial de las opciones de respuesta

Proporción				
Opción	**Total**	**Gpo. bajo**	**Gpo. alto**	**CPB**
A	0.276	0.317	0.254	-0.184
B	0.290	0.270	0.317	-0.028
C	0.157	0.143	0.143	-0.051
D*	0.263	0.238	0.270	-0.010

*Respuesta correcta; CPB=Correlación punto biserial

Fuente: Dirección de Evaluación Educativa de la CUAIEED.

La revisión de la proporción total de alumnos que seleccionó cada opción muestra que fueron seleccionadas en una cantidad semejante, con excepción de la C, confirmando la factibilidad de las alternativas de respuesta, obsérvese, sin embargo, que los distractores A y B se eligieron más que la respuesta correcta. Si se realiza una inspección por grupo de desempeño, se advierte que la proporción de alumnos del grupo alto que eligen los distractores B y C es mayor e igual, respectivamente, que los del grupo bajo, lo que confirma un mal funcionamiento de esos distractores, por otro lado, el distractor A funciona correctamente al ser mayor la proporción de alumnos del grupo bajo que lo elige. En cuanto a la respuesta correcta, aunque ciertamente la eligen más los alumnos del grupo alto, esta proporción es solo tres centésimas mayor que la del grupo bajo, por lo que en términos prácticos el reactivo no está discriminando, además el valor de la correlación punto biserial confirma el mal funcionamiento de la respuesta correcta, el resto de los distractores tiene un valor negativo en este parámetro. El análisis de los distractores invita a revisar cualitativamente este reactivo.

¿QUÉ NECESITO PARA REALIZAR UN ANÁLISIS CON TCT?

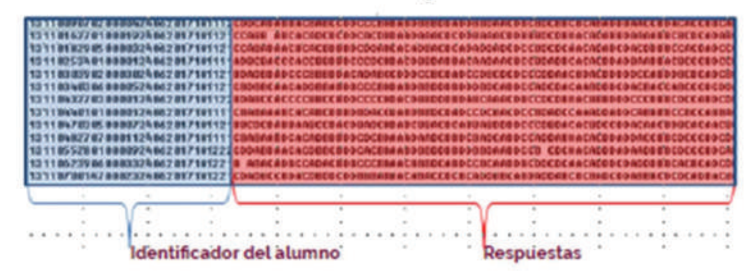
Realizar el análisis de un examen objetivo con TCT en general es un proceso ágil, ya que es factible disponer de una variedad de paquetería desarrollada para este fin, ejemplo de ello son *ITEMAN* (assess.com), *BILOG-MG* (ssicentral.com), *Jmetrik* (itemanalysis.com). En esencia, cualquiera de estos desarrollos requiere que se le alimente con las respuestas que los alumnos realizaron en el examen y la cadena de respuestas correctas, así como configurar las características del examen. A continuación, se hablará un poco del sistema desarrollado por la CUAIEED para apoyar a los interesados a realizar análisis psicométrico de exámenes objetivo.

SISAPRE

El Sistema de Análisis Psicométrico de Reactivos (SISAPRE) es un desarrollo de la CUAIEED destinado a apoyar a los interesados de cualquier parte del mundo a realizar un análisis psicométrico de un examen objetivo conforme a la TCT. Los insumos para realizar una calibración en el sistema son un archivo de texto plano, en formato dat o txt con las respuestas de los alumnos y la cadena de respuestas correctas. La Figura 2 muestra un ejemplo de los insumos.

Figura 2. Ejemplo de insumos para realizar una calibración en SISAPRE

Archivo de respuestas



Cadena de respuestas correctas

ABCDDAABCDAABCCDEBBBACCDDAAA

Fuente: Manual de usuario del SISAPRE (sisapre.cuaieed.unam.mx)

Una vez que se ingresa los insumos que solicita el sistema, el usuario puede navegar de forma amigable para analizar el comportamiento de los reactivos que conforman, para mayor detalle sobre su uso, se recomienda abrir una cuenta. Se incorporan como recursos de este capítulo tres archivos: el primero con los resultados de un examen, el segundo con la cadena de respuestas correcta y el tercero es el manual del usuario.

CONCLUSIONES Y REFLEXIONES FINALES

A lo largo de este capítulo se brindaron los principales elementos para evaluar el funcionamiento de un examen objetivo y de sus reactivos conforme a la TCT. Desarrollada en principio para brindar objetividad a evaluaciones psicológicas, su aplicación en el análisis de exámenes objetivo contribuye al diseño de evaluaciones más justas, lo cual es uno de los fundamentos que se establecen en los estándares para pruebas educativas y psicológicas (AERA, APA, NCME, 2014) a través de instrumentos más confiables.

Una de las virtudes de mayor relevancia de la TCT es la sencillez del modelo en la que se sustenta, lo que contribuye a que no sea complicada su comprensión. A pesar de su antigüedad, no pierde vigencia, prueba de ello se reporta en la segunda encuesta a profesoras y profesores de la UNAM, en el marco de los retos educativos durante la pandemia de COVID-19 (De Agüero, M. et. al., 2021) donde se exhibe que 4 de cada 10 docentes encuestados de bachillerato o licenciatura emplearon los exámenes de opción múltiple como estrategia de evaluación durante el confinamiento, lo que abre una ventana de oportunidad no solo en su diseño e implementación en línea, sino también en su evaluación, a lo que la TCT pudiera responder de una forma eficaz.

También es cierto que la TCT presenta desventajas, discutidas desde hace tiempo en trabajos como los de Gulliksen (1950), Rasch (1961), Lord & Novick (1968) y Hambleton &

Slater (1997), donde se destaca el impacto de la muestra en los resultados de la evaluación, pues esta se convierte en un sustento de validez ya sea por su tamaño como por la preservación de sus características en aplicaciones subsecuentes del instrumento. Otras limitantes que se identifican en la TCT es que, si un mismo constructo se evalúa con instrumentos distintos, los resultados no se encuentran en la misma métrica, por lo que se vuelve necesario realizar una equiparación para hacerlos comparables; por otro lado, el modelo asume que el error típico de medición se mantiene constante en todos los examinados, cuando hay individuos que pueden observar resultados más consistentes que el resto (Hambleton & Swaminathan, 1985).

Es recomendable que el docente interesado en mejorar sus instrumentos de evaluación se familiarice, como un primer paso, en los resultados que arroja la TCT y posteriormente busque incorporar otros modelos en su análisis, esto le permitirá acercarse más elementos para tomar decisiones sobre el destino de los reactivos que conforman un examen, así como evaluar la confiabilidad de su instrumento.

Actividad sugerida

En el repositorio de recursos digitales de esta obra, se tiene acceso a un archivo de texto con las respuestas de los alumnos a un examen de 54 reactivos, el cual consta de dos componentes con 27 ítems en cada uno y un identificador de cinco caracteres; un archivo con la cadena de respuestas correctas y un manual del SISAPRE. Realizar la calibración del examen y analizar la salida.

Contacto

enrique_buzo@cuaieed.unam.mx
manuel_garcia@cuaieed.unam.mx

REFERENCIAS

AERA, APA, NCME (2014). *Standars for Educational and Psychological Testing*. AERA. USA.

CUAIEED UNAM. *Manual de usuario del SISAPRE*. Recuperado de: sisapre.cuaieed.unam.mx.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297-334.

De Agüero Servín, M. Benavides, M. A. Rendón, J. Pompa, M. Hernández, A. Martínez, A. M. Sánchez, M. (2021). Los retos educativos durante la pandemia de COVID-19: segunda encuesta a profesoras y profesores de la UNAM. *Revista Digital Universitaria* (RDU), *22*(5).

DeVellis, R. F. (2006). Classical Test Theory. *Medical Care*, *44*(11), 50-59.

Gulliksen, H. (1950). *Theory of mental test*. New York: Wiley.

Hambleton, R.K. & Swaminthan, H. (1985). *Item Response Theory: Principles and applications*. Boston MA: Kluwer Nijhoff.

Hambleton, R.K. & Slater, S. C. (1997). Item Response Theory Models and Testing Practices: Current international status and future directions. *European Journal of Psychological Assessment*, *13*(1), 21-28.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Muñíz, J. (1996). *Psicometría*. Editorial Universitas, S. A. Madrid.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. I, pp. 321-334). Berkeley, CA: University of Chicago Press.

Sánchez Mendiola, M. et al. (2020). *Evaluación diagnóstica de conocimientos. Resultados de los alumnos que ingresan al bachillerato de la UNAM. 2020*. CODEIC, UNAM.

Sartes, L. M. A. Souza-Formigoni, M. L. O. (2013). Avanços na Psicometria: Da Teoria Clássica dos Testes à Teoria de Resposta ao Item. *Psicologia: Reflexão e Crítica*, *26*(2), 241-250.

Capítulo 17

UNA INTRODUCCIÓN A LA TEORÍA DE RESPUESTA AL ÍTEM PARA EL ANÁLISIS PSICOMÉTRICO DE EXÁMENES

José J. Naveja, Iwin Leenen

"Todos los modelos están equivocados, pero algunos son útiles."
George Box

INTRODUCCIÓN

En el capítulo anterior se ha abordado el enfoque de la teoría clásica de los tests (TCT) para el análisis psicométrico de los resultados de un examen. La teoría clásica está comprendida por un modelo único cuyo objetivo es estimar la "calificación verdadera" de cada uno de los sustentantes en el examen. Otro enfoque en la psicometría, más reciente, es la teoría de respuesta al ítem (TRI; en ciertas publicaciones, teoría de respuesta al reactivo). Contrario a la TCT, donde el interés está en la calificación global del examen, la TRI analiza las respuestas de los sustentantes en cada uno de los reactivos del examen, suponiendo explícitamente que estas permiten hacer una inferencia sobre una o más características psicológicas de los individuos. En la literatura psicométrica, se suele utilizar el término "rasgos latentes" para dichas características psicológicas que no son directamente observables. (En publicaciones más antiguas se hablaba de la "teoría del rasgo latente", en vez de la TRI.) Entonces, en la TRI, en contraste con la TCT, el foco se localiza en el (los) rasgo(s) latente(s) de interés y no en el instrumento que se utiliza para medirlo(s). Otra diferencia con la teoría clásica, es que la TRI incluye una gama muy amplia de modelos que permite acomodar diferentes tipos de reactivos en el examen (de respuesta abierta, de opción múltiple, etc.), así como probar distintos supuestos sobre el proceso psicológico subyacente que relaciona la respuesta en el reactivo con el (los) rasgo(s) latente(s) que se desea(n) medir.

En el presente capítulo, abordamos en primera instancia una breve reseña histórica de las ideas que dieron lugar a la TRI. Posteriormente, introducimos los principios teóricos que fundamentan el marco conceptual de esta teoría. Después, presentamos algunos ejes que permiten organizar la familia de los modelos TRI y profundizamos en algunos miembros de esta familia (sin duda, los más conocidos y utilizados para el análisis psicométrico de

exámenes). La sección subsecuente aborda ciertos temas relevantes en la aplicación práctica de los modelos TRI: estimación, bondad de ajuste y confiabilidad. Concluimos el capítulo con una discusión de las ventajas e inconvenientes del enfoque de la TRI.

ANTECEDENTES HISTÓRICOS DE LA TEORÍA DE RESPUESTA AL ÍTEM

Wim van der Linden (2016) destaca el trabajo de Alfred Binet como una contribución pionera a la TRI. A principios del siglo XX, el gobierno francés le confirió a Binet el encargo de desarrollar una prueba estandarizada para distinguir a niños con discapacidad intelectual de niños con inteligencia normal pero escasa motivación. El resultado de este proyecto fue el primer test de inteligencia, publicado por Binet en 1905 en colaboración con Théodore Simon. La innovadora metodología que Binet propuso estaba muy adelantada a su época en diferentes aspectos. En primera instancia, reconoció que la variable de interés, la inteligencia, no es directamente observable, sino latente. A Binet le interesaba desarrollar una escala para presentar sus resultados, pero concluyó que no habría una escala natural para medir un rasgo latente como la inteligencia; entonces, aplicó sus ítems a participantes de diferentes grupos etarios y asignó a cada ítem el valor de la edad cronológica con el cual el 75% de los examinados podían responderlo correctamente. Este sistema permitía estimar la edad mental de cada participante según su desempeño. Así, Binet utilizó la misma escala para medir tanto a los ítems del test como al desempeño de los participantes, lo cual es una característica fundamental de la TRI.

Dos décadas después del planteamiento de Binet, Louis Thurstone consideró necesario medir la inteligencia con su propia escala, que sería abstracta e inobservable por definición. Esto evitaría ambigüedades en la interpretación de la escala en la comparativa entre la edad mental y la cronológica: aún si la edad mental se mantuviera estable en un periodo de tiempo, el cociente intelectual disminuiría automáticamente con solo aumentar la edad cronológica. La separación de la escala del rasgo latente de otras características observables permitió a Thurstone extender los conceptos desarrollados en el contexto de la medición de la inteligencia a la medición de otras características psicológicas intangibles, como las actitudes o los rasgos de personalidad. No obstante, en el trabajo de Thurstone, y de otros investigadores que retomaron sus ideas, todavía se pueden apreciar ambigüedades en la conceptualización del rasgo latente y el uso de los modelos para su estimación.

Un paso importante en la evolución hacia la TRI consistió en utilizar funciones matemáticas para modelar la probabilidad de observar cierta respuesta en un ítem en función de la característica latente de interés en las personas. Los primeros que adoptaron estas "funciones probabilísticas de respuesta" fueron Frederic Lord (1952), con su "modelo de la ojiva normal", y George Rasch (1960), con el modelo logístico. Los avances de Rasch fueron más profundos: también propuso métodos para la estimación de parámetros y la evaluación de la bondad de ajuste del modelo. Hoy en día, el modelo de Rasch es todavía uno de los modelos más prominentes de la TRI.

Muchos otros psicómetras posteriores a Lord y Rasch hicieron avances subsecuentes que permitieron diversificar la gama disponible de modelos, por ejemplo, extendiendo los

modelos para ítems con múltiples opciones de respuesta, o incorporando más parámetros de interés, tales como la adivinación de respuestas. En la siguiente sección entraremos en los principios teóricos que comparten todos los modelos de la TRI.

PRINCIPIOS TEÓRICOS DE LOS MODELOS DE LA TEORÍA DE RESPUESTA AL ÍTEM

Todos los modelos psicométricos que se han desarrollado en el marco de la TRI comparten una serie de propiedades o principios que justamente los definen como miembros de la familia TRI. La primera propiedad es que todos estos modelos se enfocan en la probabilidad de que una persona *p*, al momento de responder un reactivo *i*, dé una respuesta en la categoría *j* del reactivo; dicha probabilidad se escribe $\Pr(Y_{pi} = j)$. Lo anterior quiere decir que un modelo TRI permite derivar, con base en sus supuestos, $\Pr(Y_{pi} = j)$. (Se utiliza el índice *p* para referirse de manera general a cualquier persona de la población para la cual se construyó la prueba; asimismo, los índices *i* y *j* se refieren de manera general a cualquier reactivo de la prueba y a cualquier categoría de respuesta en este reactivo, respectivamente. Un modelo TRI categoriza de cierta manera a las posibles respuestas; véase el inicio de la siguiente sección.)

La segunda propiedad que todos los modelos TRI comparten es la especificación de uno o más parámetros para cada persona y, de manera separada, uno o más parámetros para cada reactivo. Un parámetro es una característica teórica, ya sea de la persona o del reactivo, que se expresa a través de un valor numérico. La interpretación psicológica del parámetro depende del modelo concreto; en los modelos que asignan solo un parámetro a cada persona, este parámetro indica el nivel de la persona en el rasgo latente que se quiere medir con la prueba. En un examen de inglés, por ejemplo, el parámetro asignado a la persona *p* es un número que indica el dominio de inglés de esta persona. En la siguiente sección explicaremos para algunos modelos concretos de la TRI qué significado tienen los parámetros.

Finalmente, un modelo TRI siempre especificará una "regla" de cómo se combinan el o los parámetros de la persona *p* y el o los parámetros del reactivo *i* para llegar a la probabilidad $\Pr(Y_{pi} = j)$, comúnmente a través de una función matemática. Lo anterior se puede escribir como:

$$\Pr(Y_{pi} = j) = f\,(\text{parámetro (s) de la persona } p, \text{ parámetro(s) del ítem } i) \qquad (1)$$

En los modelos descritos en la siguiente sección, se especificará cada vez esta ecuación matemática, la cual constituye el núcleo del modelo. A continuación, se introducen algunos conceptos y términos clave de los modelos TRI: unidimensionalidad, función característica e independencia local.

Unidimensionalidad

No todos los modelos TRI son unidimensionales; sin embargo, los primeros modelos propuestos –y los más conocidos– incluyen el supuesto de unidimensionalidad. Esto quiere decir que se supone que solo hay un rasgo latente que explica las diferencias (sistemáticas) entre las respuestas de distintas personas en los reactivos de la prueba. Por ejemplo, en un examen de matemáticas, unidimensionalidad implica que solo la habilidad matemática explica, de manera sistemática, el por qué ciertos sustentantes dan mejores respuestas en el examen que otros. Aun cuando otras habilidades, como la comprensión lectora, pueden también ser relevantes o necesarias para resolver el examen (por ejemplo, para entender las instrucciones y las preguntas), un modelo unidimensional supondría que estas otras características no causan diferencias entre (las respuestas que dan) distintos sustentantes. Como mencionamos anteriormente, los modelos unidimensionales asignan solo un parámetro a cada persona (su nivel en el único rasgo latente que la prueba mide); el valor numérico de este parámetro posiciona a la persona en una "dimensión" (es decir, una línea) donde posiciones más altas corresponden con niveles más altos en el rasgo latente.

Nótese que, para muchos exámenes en la universidad, es poco plausible el supuesto de unidimensionalidad. Considérese, por ejemplo, el examen de ingreso a la UNAM; esta prueba evalúa múltiples áreas de conocimiento (español, matemáticas, historia, literatura, etcétera) y, en este sentido, es de esperar que las diferencias entre los sustentantes reveladas por el examen se deban a múltiples rasgos latentes. Incluso para un examen dedicado a un área específica, como las matemáticas, es posible que subyazcan múltiples rasgos latentes (por ejemplo, habilidad en álgebra, geometría, probabilidad, etcétera.). Si estos múltiples rasgos latentes causan diferencias entre personas y, además, son relativamente independientes (es decir, si las personas pueden tener un nivel de dominio relativamente alto en un rasgo, pero relativamente bajo en otro), entonces el supuesto de unidimensionalidad se incumple. En este caso, a veces se utilizan modelos unidimensionales como aproximación, aunque es más apropiado hacer uso de modelos multidimensionales de la TRI. Para los lectores interesados, se recomienda el libro de Reckase (2010).

Función Característica

Como se mencionó anteriormente, un modelo TRI permite calcular, a través de la ecuación básica (la cual se escribió de manera genérica en la Ecuación 1), la probabilidad de cierta respuesta con base en los parámetros de la persona y del ítem. Cuando se considera esta probabilidad como una función del (de los) rasgo(s) latente(s) implicado(s) por la prueba, se obtiene la *función característica* para esta categoría de respuesta en el ítem. En otras palabras, la función característica describe, para una categoría de respuesta en un reactivo en particular, cómo la probabilidad cambia si el nivel de la persona en el (los) rasgo(s) latente(s) cambia(n). En la frase anterior, el nivel de la persona se considera una variable (es decir, ya no hace referencia a una persona en particular) y la función característica, entonces, considera *en general* cambios en el (los) rasgo(s) latente(s) y cómo estos cambios afectan la probabilidad de una respuesta particular en un ítem específico. En la siguiente sección, donde se presentan varios modelos TRI concretos, se darán ejemplos de las funciones características.

Concluimos esta sección sobre la función característica con dos comentarios. Primero, en los modelos unidimensionales, donde solo un rasgo latente subyace a las respuestas observadas, es común representar la función característica en una gráfica, como la que se encuentra en el panel (a) de la Figura 1. En esta gráfica la abscisa (eje horizontal) indica el nivel en el rasgo latente y la ordenada (eje vertical) la probabilidad de responder en la categoría del ítem bajo consideración; el trazo describe cómo la probabilidad cambia en función del nivel en el rasgo latente. En casi todos los modelos TRI este trazo es una curva y, por eso, es común hablar de la "curva característica" como sinónimo de la "función característica".

Segundo, en los modelos para ítems dicotómicos (es decir, modelos que consideran solo dos posibles categorías de respuesta en cada ítem, por ejemplo, respuesta correcta *versus* respuesta incorrecta; véase la siguiente sección), se suele hablar de "la función característica del ítem", sin hacer referencia a una categoría de respuesta en particular. Sin embargo, la referencia es implícita y comúnmente alude a la categoría de respuesta "correcta" (más en general, la categoría que está asociada con niveles superiores en el (los) rasgo(s) latente(s). De esta manera, la función característica del ítem describe cómo *la probabilidad de acertar* el ítem cambia en función del nivel en el (los) rasgo(s) latente(s). Debe ser claro que la probabilidad de responder en la otra categoría del ítem (es decir, fallarlo) es el complemento de la probabilidad de acertar: si la probabilidad de acertar el ítem (para cierto nivel de la persona) es 0.70, entonces la probabilidad de fallarlo es 1 – 0.70 = 0.30.

Independencia Local

La ecuación básica de un modelo TRI permite calcular, para cualquier persona, su probabilidad de cierta (categoría de) respuesta para cada ítem. Por ejemplo, para el caso de exámenes de opción múltiple, hay modelos que permiten calcular la probabilidad de que la persona elija en el primer reactivo la opción B (por ejemplo, 30%), la probabilidad de que elija D en el segundo (por ejemplo, 80%), la probabilidad de que elija A en el tercero (por ejemplo, 50%), etc. Sin embargo, además de estas probabilidades para ciertas respuestas en reactivos *individuales*, los modelos TRI también especifican la probabilidad de cierto *patrón de respuestas*. En el ejemplo anterior del examen de opción, podríamos estar interesados en la probabilidad del patrón (B,D,A); es decir, la probabilidad de esta *combinación* particular de respuestas (en los primeros tres reactivos).

Para esta probabilidad, se incorpora de cierta forma un supuesto de independencia en el modelo. La forma más común consiste en suponer que una vez que se conoce el valor en el (los) parámetro(s) de la persona (es decir, su nivel en el (los) rasgo(s) latente(s) bajo estudio), conocer su respuesta en uno de los reactivos no dará información adicional acerca de qué pudo haber contestado en otro ítem. Supongamos en el ejemplo anterior que el examen de opción múltiple evalúa, como único rasgo, "dominio de inglés"; si *con base en el nivel de la persona en este rasgo* el modelo indica que la probabilidad de elegir la opción correcta en el último reactivo es igual a 50%, entonces esta probabilidad no cambiaría si conociéramos las respuestas en los reactivos previos; incluso si la persona acertara todos los reactivos previos (o los fallara todos), la probabilidad de acertar el último reactivo sigue siendo 50%.

El supuesto se llama *independencia local* (o *independencia condicional*), precisamente porque las respuestas en los otros reactivos no importan *una vez que se haya tomado en cuenta el nivel de la persona en el rasgo latente*. Con las probabilidades mencionadas en el primer párrafo de esta sección, el supuesto de independencia local implicaría que la probabilidad del patrón de respuestas (B,D,A) se obtiene multiplicando las probabilidades de los ítems individuales ($0.30 \times 0.80 \times 0.50 = 0.12$).

LA FAMILIA DE LOS MODELOS TRI

Se han desarrollado decenas (si no centenas) de modelos en el marco de la TRI. Varios autores (por ejemplo, Thissen y Steinberg, 1986) propusieron taxonomías para dar orden y estructura en esta jungla de modelos TRI. La mayoría de estas taxonomías se basan en las propiedades teóricas de los modelos. A continuación, presentamos unos ejes organizadores que sirven para diferenciar entre los distintos modelos y que se basan en criterios que son más relevantes para su aplicación:

- *Número de categorías de respuesta: modelos para ítems dicotómicos versus modelos para ítems politómicos.*

  Los primeros modelos TRI consideraban únicamente ítems dicotómicos: ítems con solo dos categorías de respuesta. Como se solían aplicar en el área de las habilidades cognitivas, estas categorías se denominaban "respuesta correcta" *versus* "respuesta incorrecta".[1] Cabe mencionar que estos modelos también se pueden utilizar cuando originalmente hay múltiples categorías de respuesta (como en los reactivos de opción múltiple de tres o más opciones), pero en estos casos se deben dicotomizar las respuestas antes de llevar a cabo el análisis, es decir, juntar ciertas opciones en la misma categoría (por ejemplo, todos los distractores de un ítem de opción múltiple se juntan en la categoría "respuesta incorrecta").

  Desarrollos posteriores permitieron analizar ítems politómicos, que en general tienen m categorías de respuesta (donde $m \geq 2$). Cuando se trata de ítems politómicos, adicionalmente se diferencia entre el caso donde las categorías estén ordenadas *a priori* (por ejemplo, cuando la calificación en los reactivos sea 0, 1, 2 o 3, y calificaciones más altas representen una mejor respuesta) *versus* el caso donde no exista un orden *a priori* (como entre las opciones de respuesta en un reactivo de opción múltiple; aunque la respuesta correcta se considera la mejor, los distractores entre sí suelen no estar ordenados). En general, analizar ítems que tienen múltiples (más de 2) respuestas con modelos para

[1] Sin embargo, esta nomenclatura no quiere decir que la aplicación de estos modelos esté limitada para el análisis de exámenes. También se pueden aplicar para analizar encuestas que miden actitudes, donde las categorías de respuesta podrían corresponder con "de acuerdo" y "en desacuerdo", o cuando se miden rasgos de personalidad, donde los reactivos son descriptores de personalidad y las categorías de respuesta son "me describe bien" o "no me describe bien".

ítems politómicos lleva a resultados más confiables en comparación con la aproximación de dicotomizar las respuestas y analizarlas con modelos para ítems dicotómicos, debido a que la dicotomización implica una pérdida de información.

- *Dimensionalidad del rasgo latente: modelos unidimensionales versus modelos multidimensionales.*
  En la sección anterior se explicó el supuesto de unidimensionalidad, que es parte de muchos de los modelos TRI. Para aplicaciones prácticas es importante considerar la plausibilidad de que solo un rasgo latente subyace a las respuestas de una prueba. En general, esta consideración guía la decisión sobre el uso de modelos unidimensionales *versus* modelos multidimensionales.

- *Forma de la función característica del ítem: modelos de dominancia versus modelos de proximidad.*
  Este capítulo se enfoca en el uso de la TRI para el análisis de exámenes. En este caso, el (los) rasgo(s) latente(s) de interés suele(n) hacer referencia a cierta(s) habilidad(es) y generalmente un nivel más alto en esta(s) habilidad(es) conlleva una probabilidad más alta de tener un mayor puntaje en el reactivo. Esto significa que la función característica (para la categoría más alta) del ítem es una función creciente. Modelos con funciones características crecientes se llaman *modelos de dominancia* (para tener un mayor puntaje en el reactivo se requiere una mayor dominancia del (los) rasgo(s) latente(s)). Efectivamente, para el análisis de exámenes se suelen utilizar modelos de dominancia.
  Sin embargo, es importante tener en mente que para cierto tipo de reactivos y cierto tipo de rasgos latentes puede ser inapropiada una función característica creciente. Consideremos, como ejemplo, el siguiente reactivo en un cuestionario para conocer la opinión de los estudiantes acerca de las medidas que se deben aplicar a profesores que acosan a las estudiantes:

  > *La sanción adecuada para profesores que acosen sexualmente a sus estudiantes es quitarles parte de su sueldo por seis meses.*

  Suponiendo que este reactivo se responde con "de acuerdo" o "en desacuerdo" y que el rasgo latente que subyace las respuestas es actitud (de rechazo) leve *versus* actitud severa hacia el acoso, es de esperar que tanto las personas con una actitud muy severa como aquellas con una actitud muy leve tengan una probabilidad baja de responder "de acuerdo" (en el primer caso porque la sanción mencionada les puede parecer insuficiente, en el segundo porque pueden considerar la sanción demasiado dura); y que personas con una actitud intermedia tendrían la probabilidad más alta de estar de acuerdo. En otras palabras, para este tipo de ítems puede ser más apropiada una función característica unimodal, es decir, una función que indica que la probabilidad aumenta hasta cierto punto y después baja. Los *modelos de proximidad* o *punto ideal* tienen funciones características

de este tipo. El lector interesado en estos modelos puede consultar los capítulos en la Sección VI en el libro de van der Linden (2016).

En el resto de esta sección, presentamos algunos de los modelos TRI más conocidos y más relevantes para el análisis de exámenes. Todos son modelos unidimensionales de dominancia. Los primeros tres son modelos para reactivos dicotómicos, con puntajes 0 (respuesta incorrecta) y 1 (respuesta correcta); el último es para reactivos politómicos con categorías ordenadas (por ejemplo, preguntas abiertas donde el puntaje mínimo es 0 y el puntaje máximo es 2 o más).

El Modelo de Rasch (o el Modelo Logístico de Un Parámetro)

Este modelo, propuesto por Rasch (1960), es uno de los más simples de los que se utilizan en la TRI. Asigna a cada persona p un parámetro, representado por θ_p, y también a cada ítem un parámetro, β_i, y combina estos parámetros (ambos números reales) a través de la siguiente ecuación para hacer una afirmación probabilística de que la persona acierte el ítem:

$$\Pr(Y_{pi} = 1) = \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)}. \tag{2a}$$

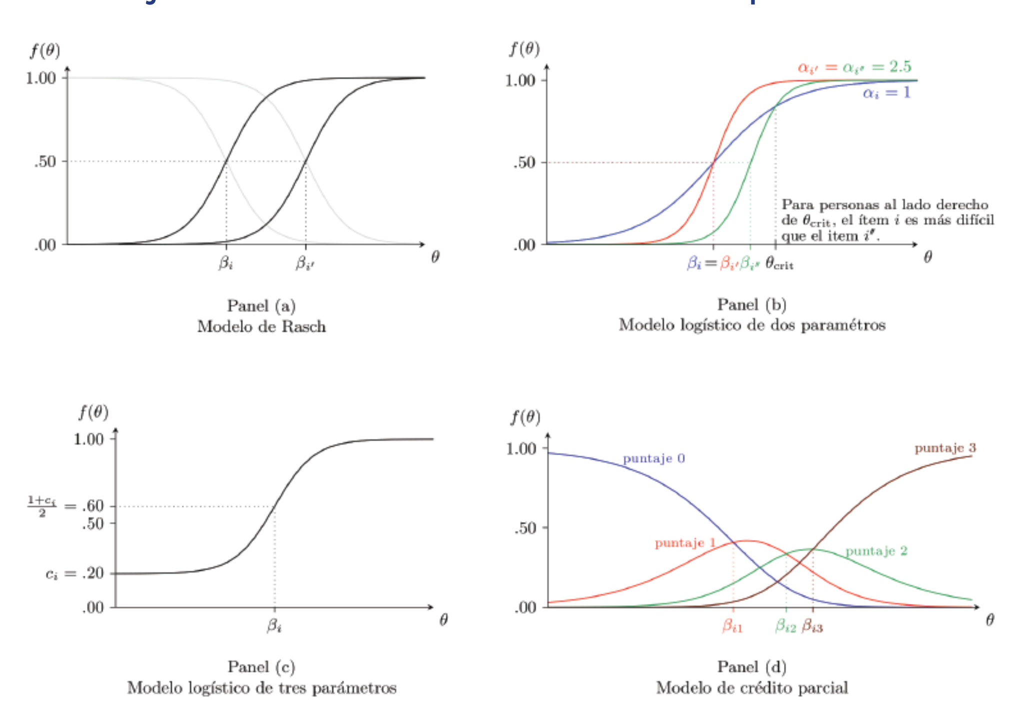
En esta ecuación $\exp(x)$ significa e^x (con e la base de los logaritmos naturales, e ≈ 2.71828). Como comentamos anteriormente, θ_p indica el nivel o el dominio de la persona p en el rasgo latente; β_i, por otro lado, se conoce como la dificultad del ítem i y corresponde con el nivel requerido en el rasgo latente para que un individuo tenga una probabilidad del 50% de dar la respuesta correcta al ítem i. (Nótese que, cuando $\theta_p = \beta_i$ en la Ecuación 2a, entonces la probabilidad de acertar el ítem es igual a 0.50.) Puesto que los ítems en el modelo de Rasch son dicotómicos, la probabilidad de que la persona falle el ítem se da por:

$$\begin{aligned} \Pr(Y_{pi} = 0) &= 1 - \Pr(Y_{pi} = 1) \\ &= 1 - \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)} \\ &= \frac{1}{1 + \exp(\theta_p - \beta_i)}. \end{aligned} \tag{2b}$$

El panel (a) de la Figura 1 muestra (en color negro) la función característica de dos ítems en el modelo de Rasch. Los ítems se colocan en la dimensión latente con base en su parámetro de dificultad, justamente –como comentamos– donde la función característica indica una

probabilidad de 50%. Nótese que el ítem i es más fácil que el ítem i' ($\beta_i < \beta_{i'}$) y que, efectivamente, para cualquier nivel de la persona en el rasgo latente, la probabilidad de acertar es mayor para el ítem i en comparación con el ítem i'. Recuérdese que la función característica de ítems dicotómicos indica la probabilidad de responder en la categoría "respuesta correcta" (dada por la Ecuación 2a); aunque normalmente no se muestra la función característica para la categoría "incorrecta", en el panel (a) de la Figura 1, incluimos para fines didácticos estas funciones características (correspondientes a la Ecuación 2b) en color gris ligero.

Figura 1. Curvas características en distintos modelos de respuesta al ítem



El Modelo Logístico de Dos Parámetros

Birnbaum (1968) añadió al modelo de Rasch un parámetro más para cada ítem. Este parámetro, comúnmente denotado α_i (un número real positivo), se suele interpretar como la discriminación del ítem i. La ecuación básica del modelo logístico de dos parámetros (2PLM, por sus siglas en inglés) es la siguiente:

$$\Pr(Y_{pi} = 1) = \frac{\exp[\alpha_i(\theta_p - \beta_i)]}{1 + \exp[\alpha_i(\theta_p - \beta_i)]}. \tag{3}$$

Nótese que, si el parámetro de discriminación α_i es igual a uno para todos los ítems, el modelo 2PL se simplifica y se obtiene el modelo de Rasch. El panel (b) de la Figura 1 mues-

tra la función característica de tres ítems en el 2PLM. Se observa que los ítems con un grado de discriminación más alto (por ejemplo, el ítem de color más oscuro) tienen la pendiente más pronunciada. Efectivamente, es lo que significa la discriminación de un ítem en este modelo: para dos personas (con posiciones cercanas al grado de dificultad del ítem) se hace más grande la diferencia entre sus probabilidades de acertar conforme el ítem discrimina más. En el caso más extremo (cuando $\alpha_i \to \infty$), la discriminación es máxima: personas con un nivel de habilidad menor que la dificultad β_i del ítem tienen probabilidad 0 de acertar, mientras que los que tienen el nivel arriba de β_i tienen probabilidad 1.

La gráfica muestra adicionalmente una diferencia importante entre el 2PLM y el modelo de Rasch: las curvas características en el 2PLM generalmente se cruzan, lo cual nunca puede ocurrir en el modelo de Rasch. Como consecuencia de esta característica, la interpretación del parámetro de dificultad ya no es uniforme para todas las personas. En la gráfica se observa, por ejemplo, que $\beta_i < \beta_{i''}$, mientras que para las personas con un nivel mayor que θ_{crit} (indicado en la gráfica), el ítem i'' es más fácil que el ítem i. En general, en modelos TRI más complejos (con más parámetros), la interpretación de los parámetros suele ser menos unívoca.

La gráfica muestra también una similitud entre el modelo de Rasch y el 2PLM: para personas con un nivel θ_p muy bajo en el rasgo latente, la probabilidad de acertar el ítem se acerca a 0; en contraste, las personas con la θ_p muy alta tienen probabilidades cercanas a 1. El siguiente modelo permite curvas características donde la probabilidad no se acerca a 0 para niveles muy bajos en el constructo latente.

El Modelo Logístico de Tres Parámetros

Los modelos anteriores pueden presentar una inconveniencia cuando se utilizan para analizar exámenes de opción múltiple. En un examen de este tipo es poco plausible que la probabilidad de acertar se acerque a 0, incluso para personas con un dominio muy bajo en el rasgo latente, ya que se puede responder correctamente por adivinación (azar). Es por esta razón que Birnbaum (1968) propuso el modelo logístico de tres parámetros (3PLM). En este modelo, además de los parámetros α_i y β_i, cada ítem tiene un tercer parámetro, c_i, el cual se denomina "parámetro de adivinación". Este parámetro (cuyo valor teóricamente puede estar entre 0 y 1) se suele interpretar como la probabilidad de acertar el ítem en caso de que una persona no sepa la respuesta. La ecuación básica que permite calcular la probabilidad de acertar en el 3PLM se da por:

$$\Pr(Y_{pi} = 1) = c_i + (1 - c_i) \frac{\exp[\alpha_i(\theta_p - \beta_i)]}{1 + \exp[\alpha_i(\theta_p - \beta_i)]}.$$

Es claro que este modelo generaliza el 2PLM: cuando en la ecuación anterior $c_i = 0$ para todos los ítems, se obtiene la Ecuación 3. El panel (c) de la Figura 1 muestra como ejemplo

la curva característica de un reactivo en el 3PLM. Se observa que, efectivamente, la curva no baja hasta una probabilidad de 0 para las personas con el nivel θ_p muy bajo, sino a la probabilidad de $c_i = 0.20$ (en este ejemplo). Además, se observa que el parámetro β_i, contrario a los modelos anteriores, ya no corresponde con el nivel en el rasgo latente donde la probabilidad de acertar el ítem es igual a 50% (sino donde la probabilidad es igual a $(1 + c_i) / 2$).

Para una interpretación correcta del parámetro c_i en el 3PLM son útiles las siguientes dos reflexiones. Primero, cada ítem tiene un parámetro c_i y distintos ítems suelen tener distintos valores para c_i. A veces surge una confusión porque, de ser interpretado como la probabilidad de elegir por azar la opción correcta en un reactivo de opción múltiple, el parámetro c_i debería tener necesariamente el valor $1/m$, donde m es el número de opciones de respuesta para el reactivo (por ejemplo, $c_i = 0.25$ si el reactivo tiene cuatro opciones de respuesta). Si bien este razonamiento sería correcto si una persona respondiera el reactivo *ciegamente* (como si no hubiera leído las opciones), el 3PLM toma en cuenta que la opción correcta puede ser más o menos atractiva para personas con un nivel muy bajo en el rasgo latente, por ejemplo, porque estas personas se dejan atraer más por algún distractor o, al contrario, algún distractor es obviamente incorrecto, incluso para personas "ignorantes". Segundo, es importante tener claro que c_i es un parámetro asociado con los ítems y, en este sentido, hace referencia a una propiedad de los ítems; esto implica que, si cierto reactivo de opción múltiple tiene un valor alto para c_i, entonces la opción correcta de este reactivo es relativamente atractiva para *todas* las personas con nivel bajo en el rasgo latente. Específicamente con respecto a los exámenes de opción múltiple se ha introducido el rasgo de *testwiseness*, lo cual se define como la habilidad que pueden tener ciertas personas para aprovechar aspectos del formato y de la redacción de los reactivos para aumentar su calificación en el examen. Sin embargo, esta habilidad, al ser una característica de las personas (ajena al constructo que con el examen se quiere medir), no puede ser captada por el parámetro c_i en el 3PLM. (Se podría tomar en cuenta el rasgo de *testwiseness* en un modelo TRI; este sería necesariamente un modelo multidimensional, ya que, además del rasgo latente que se quiere medir, incluiría también el *testwiseness*.)

El Modelo de Crédito Parcial

Si las respuestas en las preguntas de un examen no son binarias (correcto-incorrecto) o no se desea dicotomizarlas para que así lo sean, los tres modelos previos no sirven para su análisis. En este caso hay que recurrir a modelos TRI para ítems politómicos. Aquí describimos brevemente uno de estos modelos: el modelo de crédito parcial (PCM, por sus siglas en inglés) de Masters (1982, 2016).

El PCM considera, de manera general, el caso de un examen donde cada pregunta se califica con un número entero de 0 a m_i (con $m_i \geq 1$; se permite que la puntuación máxima es diferente para distintos ítems, por eso se añadió el índice i en m_i), donde una puntuación mayor es indicador de un nivel más alto en el (único) rasgo latente que el examen quiere medir. Por un lado, este modelo, igual que los anteriores, asigna un parámetro, θ_p, a cada persona; por otro lado, a cada ítem se asignan m_i parámetros (un parámetro para cada

categoría/puntaje mayor que 0): β_{i1}, β_{i2} , ..., β_{im}. La ecuación básica, que no presentamos aquí por su complejidad matemática (pero se puede consultar en las publicaciones mencionadas de Masters), permite calcular, con base en estos parámetros, la probabilidad de que la persona p en el ítem i obtenga un puntaje igual a j (para cualquier valor de $j = 0,1, ..., m_i$).

El panel (d) de la Figura 1 muestra un ejemplo de un ítem en el PCM. Debido a que se trata de un ítem politómico con puntaje máximo de $m_i = 3$, la gráfica muestra cuatro curvas características distintas (asociadas con los puntajes 0, 1, 2 y 3). Nótese que las curvas características asociadas con el puntaje mínimo de 0 y el puntaje máximo de 3 son monótonamente decrecientes y crecientes, respectivamente, mientras que las categorías intermedias tienen curvas unimodales.[2] Esta forma de las curvas no sorprende: efectivamente, conforme el nivel de la persona en el rasgo latente aumenta, la probabilidad del puntaje de 0 baja y la del puntaje máximo crece. Por otro lado, con respecto a los puntajes intermedios, ni las personas con un nivel extremadamente bajo ni con un nivel extremadamente alto tendrán estos puntajes; las primeras porque no logran superar el puntaje de 0, las últimas porque logran tener el puntaje máximo de 3.

La gráfica nos muestra también cómo se pueden interpretar los parámetros β_{i1}, β_{i2} y β_{i3} de este ítem (y, en general, de los ítems en el PCM). Se observa que β_{ij} corresponde con la posición en el rasgo latente donde las curvas características asociadas con los puntajes j y $j - 1$ se cruzan. Por ejemplo, β_{i1} indica el nivel en el rasgo latente desde el cual es más probable tener un puntaje de 1 que un puntaje de 0, y en este sentido se puede interpretar como la dificultad de la categoría 1 relativa a la categoría 0. Nótese que la misma interpretación aplica al único parámetro (β_i) del ítem en el modelo de Rasch, que indica el nivel en el rasgo latente donde acertar y fallar tienen la misma probabilidad. En efecto, el PCM es una generalización para ítems politómicos del modelo de Rasch y cuando $m_i = 1$, el PCM es idéntico al modelo de Rasch. Cabe mencionar que en desarrollos posteriores del PCM, se incluyó también un parámetro de discriminación (α_i) para cada ítem (similar al 2PLM), lo cual llevó al modelo de crédito parcial generalizado.

CONSIDERACIONES RELEVANTES EN LA APLICACIÓN DE UN MODELO TRI

Hasta este punto, hemos dado una introducción teórica a los modelos TRI. A fin y al cabo, los modelos estadísticos son una teoría, formalizada en lenguaje matemático, para describir un fenómeno; en el caso de los modelos psicométricos de la TRI, son una teoría sobre la conducta de personas que se enfrentan a (los reactivos en) un examen. Sin embargo, para que los

2 Al inicio de esta sección, asociamos las curvas características unimodales con los modelos de proximidad o punto ideal. Para una comprensión correcta, estos modelos especifican curvas características unimodales, no solo para las posibles categorías intermedias (en caso de ítems politómicos), sino también para la categoría más alta. El PCM no es un modelo de proximidad (sino de dominancia) ya que la curva característica asociada con el puntaje máximo es creciente (o, de manera equivalente: conforme el nivel en el rasgo latente aumenta, el puntaje esperado en el ítem aumenta también).

modelos tengan relevancia práctica es importante concretar su aplicación a datos observados. En general, tal aplicación implica resolver las siguientes tres cuestiones: (1) ¿Qué estrategia se utilizará para asignar valores a los parámetros incógnitos del modelo? Este es el problema que resuelve la estimación de parámetros. (2) Una vez que se han estimado los valores de los parámetros, ¿se puede concluir que el modelo da una descripción adecuada de los datos observados? Esta cuestión se refiere a la bondad de ajuste del modelo. (3) Tomando en cuenta que en las aplicaciones prácticas siempre hay factores no deseados que distorsionan los datos observados, ¿qué tan confiables son los resultados (es decir, las estimaciones de los parámetros) obtenidos mediante el modelo? Esta cuestión hace referencia a la precisión de las estimaciones. A continuación, exponemos brevemente cómo estas cuestiones se suelen abordar en el contexto de los modelos TRI.

Estimación de Parámetros

En el caso de la TRI, los datos observados a los que se aplican los modelos son las respuestas de una muestra de personas al conjunto de reactivos que conforman el examen. En la sección anterior presentamos los principios generales de los modelos TRI y explicamos de manera breve algunos de estos modelos que son muy utilizados en la práctica. Aclaramos que los modelos especifican parámetros (valores numéricos con un significado particular en el contexto de un modelo concreto), para personas e ítems, pero no mencionamos cómo se obtienen estos valores numéricos. Un primer paso en la aplicación de un modelo TRI es el ajuste del modelo a datos observados, es decir, la estimación de sus parámetros con base en estos datos.

Consideremos como ejemplo una aplicación del modelo de Rasch a los datos que se presentan (parcialmente) en la Tabla 1; son las respuestas (1 = correcto; 0 = incorrecto) de 500 personas a 60 reactivos. Aplicar el modelo de Rasch a estos datos implica estimar los 500 parámetros de personas y los 60 parámetros de ítems. En la Tabla 1 hemos añadido el acento ^ al parámetro (por ejemplo: $\hat{\theta}_p$ y $\hat{\beta}_i$) en la última columna y fila, para diferenciar el valor estimado con base en datos observados del valor teórico-verdadero en el modelo.

Existen diferentes métodos para estimar los parámetros de un modelo estadístico. Para los modelos TRI, el método más utilizado es la estimación por máxima verosimilitud (MLE, por sus siglas en inglés) y es la estrategia comúnmente implementada en los algoritmos de estimación en *software* comercial como BILOG, MULTILOG, PARSCALE (*Scientific Software International Inc.*) y WinSteps (*Winsteps Inc.*) y *software* libre como MIRT (*Educational Testing Service*), jMetrik (Universidad de Virginia) y los paquetes especializados de R. En el marco de este capítulo introductorio no tenemos el espacio para indagar sobre la MLE. Solo mencionamos que existen distintas variantes de MLE y que cada variante tiene sus ventajas y desventajas: el método por máxima verosimilitud marginal es muy flexible, pero añade un supuesto al modelo (el supuesto de que los parámetros θ_p se extraen de manera aleatoria de una distribución normal); el método por máxima verosimilitud condicional no requiere supuestos adicionales, pero solo es aplicable a un subgrupo de modelos (que incluyen la propiedad de que el puntaje total de una persona en el test es suficiente para estimar el

parámetro θ_p de esta persona); y el método de máxima verosimilitud conjunta, aunque generalmente aplicable, tiene la desventaja teórica de que las estimaciones no son consistentes. Cabe mencionar que, aunque (las variantes de) MLE sigue siendo el procedimiento estándar de estimación para modelos TRI, cada vez es más común realizar estimaciones en el marco bayesiano.

Evaluación de la Bondad de Ajuste del Modelo

Los métodos de estimación asignan valores numéricos a los parámetros de personas e ítems que, en cierto sentido, son los mejores. Sin embargo, incluso con las mejores estimaciones puede ser que el modelo no describa de manera aceptable los datos observados; en concreto, puede ser que las probabilidades derivadas a partir de la ecuación básica del modelo no tengan una correspondencia adecuada con las respuestas de las personas a los reactivos. Por ejemplo, considérese en una aplicación del modelo de Rasch a una persona cuyo nivel en el rasgo latente se estimó como intermedio, mientras que los datos muestran que esta persona acierta los ítems difíciles (con altos valores para $\hat{\beta}_i$) y falla los ítems fáciles (con valores bajos para $\hat{\beta}_i$). Esta situación señala una discrepancia entre los datos y el modelo, ya que las probabilidades de acertar según el modelo son altas justamente donde la persona falla y, al contrario, son bajas cuando la persona acierta. Aunque el modelo no excluye la posibilidad de que ocurran estas discrepancias, si en los datos ocurren frecuentemente, lo más apropiado entonces sería rechazar el modelo y buscar otro modelo TRI que dé una mejor explicación a los datos observados.

Un mal ajuste del modelo a los datos apunta a que uno o más supuestos se están incumpliendo. Existen muchas estrategias para evaluar la bondad de ajuste de un modelo TRI. Estas estrategias pueden evaluar el ajuste de manera global o bien, pueden detectar incumplimientos particulares, quizás asociados con ciertas personas o ítems, o con supuestos específicos (como unidimensionalidad o independencia local). A continuación, describimos brevemente los principios de dos grupos de métodos de evaluación de la bondad de ajuste para modelos TRI.

- *Métodos gráficos/visuales*
  Este método (más útil para modelos unidimensionales) evalúa la bondad de ajuste de los distintos ítems a través de una gráfica que permite comparar (las probabilidades de responder en cierta categoría mostradas por) la curva característica de cada ítem con la proporción de respuestas en distintos subgrupos de personas que tienen un nivel estimado similar en el rasgo latente. Para ilustrarlo, volvamos a considerar los datos en la Tabla 1 y la aplicación del modelo de Rasch a estos datos. Como siguiente paso, dividimos el grupo total de 500 personas en 10 grupos de aproximadamente 50 personas que tienen estimaciones $\hat{\theta}_p$ similares (es decir, un primer grupo con las 50 personas con las $\hat{\theta}_p$ más bajas, después un segundo grupo con las siguientes 50 personas, etcétera.) y calculamos en cada grupo la proporción de aciertos en el ítem i bajo consideración. En la gráfica de la curva característica se añade un punto a la altura de la proporción calculada

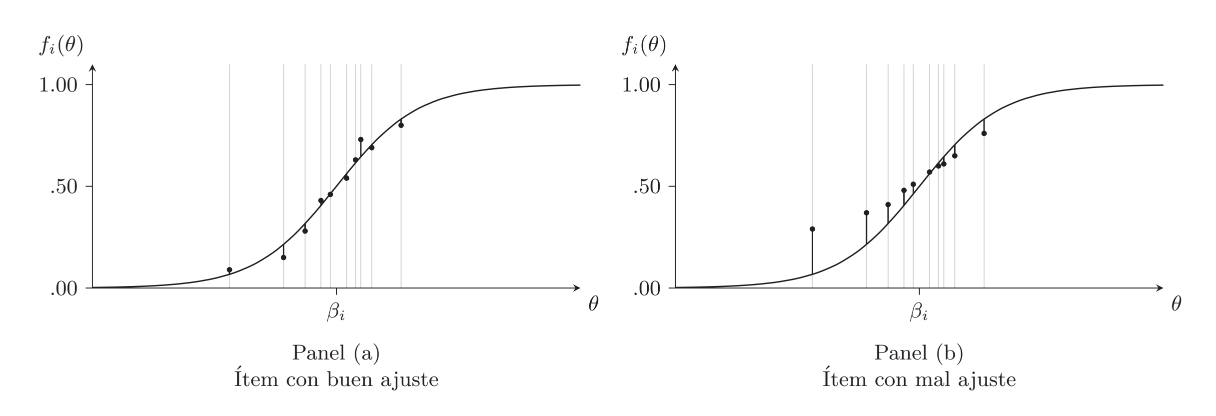
para cada grupo (donde el grupo se posiciona en la abscisa según la media de las $\hat{\theta}_p$). El panel (a) de la Figura 2 muestra un ejemplo de un ítem con buen ajuste: las proporciones de aciertos en los distintos grupos se acercan a (las probabilidades teóricas en) la curva; el panel (b) muestro un ítem con mal ajuste, donde las discrepancias entre las proporciones y las probabilidades son más grandes.

Tabla 1. Ejemplo de datos observados y los parámetros asociados en el modelo de Rasch.

		Ítems									
		1	2	3	4	$\cdots$	i	$\cdots$	60		
Personas	1	1	0	0	1	$\cdots$	1	$\cdots$	1	$\rightarrow$	$\hat{\theta}_1$
	2	0	1	0	1	$\cdots$	1	$\cdots$	0	$\rightarrow$	$\hat{\theta}_2$
	3	0	0	1	1	$\cdots$	1	$\cdots$	0	$\rightarrow$	$\hat{\theta}_3$
	4	0	1	0	0	$\cdots$	0	$\cdots$	0	$\rightarrow$	$\hat{\theta}_4$
	5	0	0	0	1	$\cdots$	0	$\cdots$	1	$\rightarrow$	$\hat{\theta}_5$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$
	p	1	1	0	0	$\cdots$	1	$\cdots$	1	$\rightarrow$	$\hat{\theta}_p$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$		$\vdots$
	500	1	0	0	1	$\cdots$	0	$\cdots$	0	$\rightarrow$	$\hat{\theta}_{500}$
		$\downarrow$ $\hat{\beta}_1$	$\downarrow$ $\hat{\beta}_2$	$\downarrow$ $\hat{\beta}_3$	$\downarrow$ $\hat{\beta}_4$	$\cdots$	$\downarrow$ $\hat{\beta}_i$	$\cdots$	$\downarrow$ $\hat{\beta}_{60}$		

Nota. La última columna y la última fila indican los parámetros estimados de personas e ítems, respectivamente, al aplicar el modelo de Rasch a estos datos.

Figura 2. Evaluación de la bondad de ajuste de dos ítems en el modelo Rasch (método gráfico)



- *Índices de bondad de ajuste*
  Los programas informáticos para aplicar modelos TRI permiten calcular una amplia gama de índices de bondad de ajuste. Estos índices cuantifican de cierta manera la diferencia entre los datos observados y los datos esperados según el modelo (similar al ejemplo que se acaba de describir en el método gráfico, donde de manera visual se aprecia la diferencia entre las proporciones de aciertos observadas en la muestra y las probabilidades teóricas, que se pueden interpretar como proporciones esperadas según el modelo). Comúnmente se distingue entre índices exactos, que permiten realizar una prueba exacta de bondad de ajuste del modelo a los datos (donde un ejemplo clásico es el estadístico ji-cuadrado) e índices de ajuste aproximados. Estos últimos se construyeron a partir de la idea que cualquier modelo es equivocado en el sentido de que no puede abstraer la realidad con todos sus detalles y, por lo tanto, se rechazará cuando hay suficientes datos disponibles. Sin embargo, al proponer un modelo se espera poder usarlo para obtener información y conocimiento conciso acerca de los datos observados y es suficiente con que el modelo tenga un ajuste suficientemente bueno, aunque no sea perfecto. Un índice de bondad de ajuste aproximado muy utilizado es el RMSEA, que con valores más bajos indica un mejor ajuste. Para este tipo de índices, la literatura presenta puntos de corte a partir de los cuales se considera que el ajuste es "aceptable" o "bueno" (por ejemplo, RMSEA $< .08$ se considera un ajuste aceptable; RMSEA $< .05$ es excelente).

Antes de concluir esta sección sobre bondad de ajuste, queremos llamar la atención a un incumplimiento de los supuestos en modelos TRI que en los últimos años ha recibido más atención y cuya detección en exámenes de alto impacto se considera esencial. Se trata de *funcionamiento diferencial del ítem* (DIF). Uno de los supuestos clave en los modelos TRI es que los parámetros de los ítems tengan el mismo valor (técnicamente se dice que son "invariantes") para cualquier muestra de personas extraída de la población para la cual el modelo se cumple (y, de forma similar, que los parámetros de las personas son invariantes para dis-

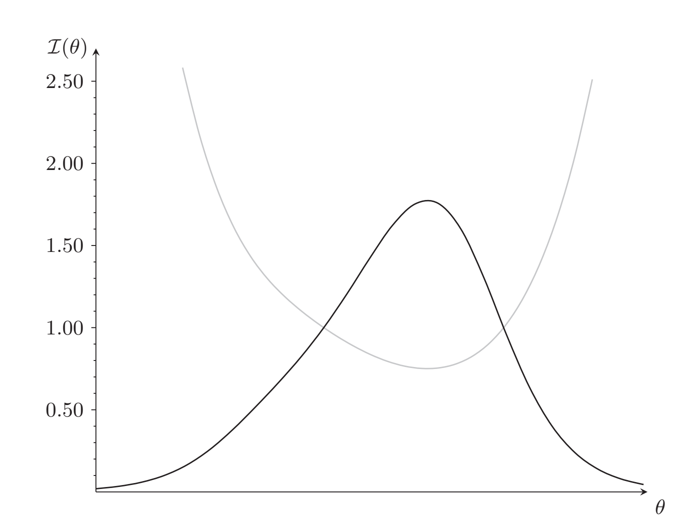
tintas submuestras de ítems). Se presenta DIF si los ítems se comportan de manera diferente en distintas subpoblaciones, por ejemplo, con estudiantes de familias con recursos limitados o en la subpoblación de mujeres, etc. Cuando se presenta funcionamiento diferencial, las comparaciones de distintas subpoblaciones pueden resultar inapropiadas (en el sentido de que una $\hat{\theta}_p$ más baja en un grupo no necesariamente corresponde con un nivel más bajo en el rasgo latente que se quiere medir). Existe mucha literatura que presenta métodos para evaluar y detectar DIF en pruebas analizadas con modelos TRI.

Confiabilidad (Precisión de las Estimaciones)

En la sección sobre estimación de parámetros, enfatizamos la diferencia conceptual entre los valores teóricos para los parámetros (θ_p, β_i, α_i, etc.) y las estimaciones correspondientes ($\hat{\theta}_p$, $\hat{\beta}_i$, $\hat{\alpha}_i$, etc.). Es de esperar que, por la influencia de factores aleatorios, los valores estimados no coincidan perfectamente con los teóricos (los cuales, en principio, son siempre desconocidos). De ahí surge el tema de qué tanto difieren las estimaciones de los valores verdaderos. Aunque nunca sabremos para un caso concreto cuál es la diferencia entre, digamos, θ_p y $\hat{\theta}_p$, sí es posible hacer afirmaciones generales sobre la precisión de las estimaciones (por ejemplo, cuál es la diferencia esperada –o promedio– entre la estimación y el valor teórico-verdadero correspondiente).

En la teoría clásica de los tests, la precisión de la medición depende de la confiabilidad del test. En particular, la confiabilidad permite calcular el error estándar de medición (el cual es la desviación estándar del error de medición, y en este sentido, se interpreta como la diferencia esperada entre la puntuación verdadera y la puntuación observada en un test). En la TRI, se toma otro enfoque con respecto al error de medición, el cual dependerá de la estrategia de estimación particular que se utilice. En el caso de MLE, por ejemplo, la precisión de la medición (es decir, la diferencia esperada entre el valor estimado y el valor teórico-verdadero del parámetro) se cuantifica a través de la función de información de Fisher. Cuando se aplica a la estimación del parámetro θ_p, la información de Fisher tiene un papel similar a la confiabilidad en la TCT, en el sentido de que permite calcular el error estándar de medición (en particular, el error estándar de medición es el inverso de la raíz cuadrada de la información). Una diferencia fundamental entre la información de Fisher y la confiabilidad (o entre el error estándar de medición en la TRI *versus* la TCT) es que la información depende (del valor del parámetro) de la persona (véase la Figura 3), mientras que la confiabilidad es constante en una población de personas. Es decir, en la TRI las $\hat{\theta}_p$ tienen una precisión distinta para diferentes personas que contestan el test. Intuitivamente, no resulta extraño que la precisión de las estimaciones del nivel en el rasgo latente sea baja para personas con calificaciones muy altas o muy bajas. (Si necesitáramos una precisión más alta para discriminar entre sustentantes en los extremos del rasgo latente, sería oportuno realizar otras pruebas, por ejemplo, un test adaptativo que ajuste automáticamente el nivel de dificultad de los ítems que se presentan al sustentante conforme avanza en la prueba.) Cabe mencionar que también en un marco bayesiano existen métodos para cuantificar la precisión de las estimaciones, la cual puede variar entre distintas personas.

Figura 3. Ejemplo de la función de información (negro) y el error estándar de medición (gris)



DISCUSIÓN Y CONCLUSIONES

En este capítulo hemos dado una introducción a la TRI dirigida a profesores sin antecedentes técnicos en psicometría, con una descripción breve de algunos de los modelos más utilizados y temas relevantes para su aplicación en el análisis de exámenes. Debido a sus fundamentos arraigados en la estadística matemática, publicaciones sobre la TRI suelen estar intercaladas con fórmulas complejas y lenguaje técnico difíciles de entender por usuarios no expertos en este tema. Esto indudablemente es la razón principal por la que el uso de modelos TRI para el análisis de exámenes, a pesar de sus numerosas ventajas sobre la TCT, sigue siendo relativamente marginal.

Otro inconveniente de la TRI, que muchos autores mencionan para explicar su uso poco frecuente en la práctica, es la necesidad de muestras relativamente grandes para la aplicación de los modelos. Incluso para modelos relativamente sencillos (como los presentados en este capítulo) se requieren como mínimo unos cuantos cientos de personas para obtener estimaciones suficientemente precisas. En este sentido, es complicado aplicar los modelos TRI para exámenes que se aplican a grupos de 60 personas, como es el caso para muchos profesores en instituciones educativas.

Concluimos este capítulo mencionando dos de las ventajas más importantes de la TRI sobre la TCT. Aunque más de índole teórico, los autores consideramos que la ventaja principal de la TRI es que da un fundamento robusto para la medición del (los) constructo(s) latente(s) que se quiere(n) medir con el examen. Mientras que los modelos TRI describen explícitamente (de manera probabilística, a través de la ecuación básica del modelo) cómo las respuestas en el examen dependen del nivel de la persona en el (los) constructo(s) latente(s), la teoría clásica ni siquiera hace referencia a un constructo latente en el desarrollo del modelo. (Nótese, al respecto, que la puntuación verdadera en la TCT no solo recoge el efecto

del (los) rasgo(s) latente(s) que se quiere medir con el test, sino también de *todos* los factores que ejercen una influencia sistemática entre las distintas réplicas conceptuales en las que se basa la definición de la puntuación verdadera y del error.) La TRI permite llegar a conclusiones sobre la dimensionalidad del rasgo latente (¿cuántos rasgos latentes subyacen al test y cómo se relacionan entre sí?) y, además, cómo esta(s) dimensión(es) se relaciona(n) con los reactivos en el examen.

La segunda ventaja importante se deriva de la propiedad, anteriormente mencionada, de la invarianza de los parámetros del modelo. Aunque son importantes las consideraciones sobre el diseño para la recopilación de los datos y el método de estimación (véase Hambleton, 1994), esta invarianza generalmente ofrece una base sólida para comparar los parámetros de diferentes personas, aunque respondieran distintos conjuntos de ítems y, viceversa, para comparar los parámetros de diferentes ítems, aunque hayan sido respondidos por distintos grupos de personas. Esta propiedad conlleva ventajas importantes cuando, es imposible (o al menos no deseable) aplicar el mismo examen a todas las personas, por ejemplo, cuando un examen se aplica en diferentes días a diferentes grupos y existe el riesgo que los reactivos sean compartidos entre los sustentantes. En estos casos, un análisis con base en la TCT no permitiría diferenciar entre puntuaciones más altas porque la variante del examen aplicada en cierto día era más fácil o bien porque los sustentantes de ese día tenían un nivel más alto en el (los) rasgo(s) evaluado(s). Cabe mencionar que la explotación de la propiedad de invarianza se lleva a un extremo en el caso de los llamados "tests adaptativos informatizados", que presentan, a través de una computadora, a cada persona un conjunto de reactivos personalizado mientras que los niveles estimados de distintas personas en el (los) rasgo(s) latente(s) son directamente comparables. El uso de tests adaptativos informatizados lleva a exámenes más cortos y estimaciones más precisas, aún para sustentantes con valores extremos en el (los) rasgo(s) latente(s), debido a que cada persona responde a ítems ajustados a su nivel. Este tipo de pruebas, que hacen uso exquisito de la tecnología, no son compatibles con la TCT, pero sí con la TRI.

Agradecimiento

Los autores agradecen a Georgina García Rodríguez por la lectura crítica y comentarios realizados sobre una versión previa de este capítulo.

REFERENCIAS

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. En F. M. Lord y M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 396–479). Addison-Wesley.

Hambleton, R. K. (1994). Item response theory: A broad psychometric framework for measurement advances. *Psicothema*, *6*(3), 535–556.

Lord, F. M. (1952). *A theory of test scores* (Psychometric Monograph No. 7). Psychometric Corporation. https://www.psychometricsociety.org/sites/main/files/file-attachments/mn07.pdf

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174.

Masters, G. N. (2016). Partial credit model. En W. J. van der Linden (Ed.). *Handbook of item response theory* (Vol. 1, pp. 109–126). CRC Press, Taylor & Francis.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Danish Institute for Educational Research.

Reckase, M. D. (2009). *Multidimensional item response theory.* Springer.

Thissen, D., y Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, *51*(4), 567–577.

van der Linden, W. J. (Ed.). (2016). *Handbook of item response theory* (Vol. 1). CRC Press, Taylor & Francis.

Capítulo 18

UNA INTRODUCCIÓN A LA TEORÍA DE LA GENERALIZABILIDAD

Iwin Leenen, Olivia Espinosa Vázquez

"No he fracasado, he encontrado 10,000 formas que no funcionan."
THOMAS A. EDISON

INTRODUCCIÓN

Como se señaló en el capítulo 16 de este libro, la teoría clásica de los tests (TCT) se fundamenta en una base axiomática comprendida en la ecuación $X = T + E$, donde X y T son variables que representan a la puntuación observada y la puntuación verdadera, respectivamente, asociadas con una medición particular considerada para una población de sustentantes. Por otro lado, E representa el error asociado con la medición y técnicamente "explica" por qué la puntuación observada de un sustentante no coincide perfectamente con su puntuación verdadera. (Nótese que de la ecuación básica resulta que $E = X - T$.) A partir del supuesto de que la correlación entre T y E es igual a 0 (o, más general, que el error de medición es independiente de la puntuación verdadera), se deriva de la ecuación básica el siguiente resultado fundamental:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2. \tag{1}$$

En palabras, la varianza observada (es decir, las diferencias entre sustentantes que la medición revela) se parte en dos: varianza verdadera (diferencias entre sustentantes que se deben a diferencias en sus puntuaciones verdaderas) y varianza error (diferencias atribuibles a factores que perturban la medición de una manera no sistemática). La Ecuación 1 directamente lleva a la definición de confiabilidad en el marco de la TCT:

$$\rho_{XX'} = \frac{\sigma_T^2}{\sigma_X^2}\ , \tag{2}$$

es decir, la confiabilidad es la proporción de la varianza observada atribuible a la puntuación verdadera.

Desde su concepción en la primera mitad del siglo XX, la TCT ha permanecido como el modelo más utilizado para construir y evaluar instrumentos psicológicos. En este capítulo presentamos una extensión a la TCT, que se conoce como la Teoría de la Generalizabilidad (Teoría G), que tiene sus origines en las publicaciones de Lee Cronbach y colegas (1963, 1972). La Teoría G se diferencia de la TCT en el sentido de que distingue entre múltiples fuentes de error y, de ahí, por una manera diferente (a la presentada en la Ecuación 1) de particionar la varianza observada. Un objetivo esencial de la Teoría G consiste precisamente en estimar la contribución de las distintas fuentes que influyen en la medición, es decir, la importancia que tienen en la varianza de las puntuaciones observadas.

La estructura del capítulo es como sigue: en la siguiente sección introducimos dos casos que representan, cada uno, una evaluación típica en un contexto universitario y que utilizaremos como guía para ilustrar los conceptos que se desarrollan en el resto de este capítulo. Después, presentamos el marco conceptual de la Teoría G, con sus definiciones más importantes. Puesto que la manera en que se concreta la Teoría G depende de las características y el diseño de la evaluación específica a la que se aplica, las siguientes secciones elaboran la Teoría G en el contexto de los dos casos introducidos en la siguiente sección, que llevan a desarrollos relativamente sencillos en el marco de la Teoría G. En la sección subsiguiente describimos de una manera muy general aspectos de diseños avanzados, más complejos. Concluimos el capítulo con algunos comentarios sobre relaciones entre la Teoría G y otros enfoques psicométricos.

EJEMPLOS DE GUÍA

A continuación, describimos dos ejemplos de exámenes típicos en un contexto universitario. El examen A evalúa a los estudiantes de la materia de Cálculo Diferencial e Integral de la carrera de Matemáticas. Participan 68 estudiantes en el examen, el cual consiste en 20 preguntas que son integrales definidas que el sustentante tiene que resolver. Cada pregunta se califica únicamente considerando los puntajes de 0 y 1 (respuesta incorrecta y correcta, respectivamente). La calificación en el examen es la proporción (o el porcentaje) de las 20 integrales que el sustentante resuelve correctamente. Los datos de este examen se pueden estructurar como se muestra en la parte superior de la Tabla 1.

Tabla 1. Estructura de los datos observados para los Exámenes A y B

Examen A

Sustentantes (*p*)	Preguntas (*i*) 1	2	3	4	5	⋯	19	20	$\bar{X}_{pI}$
1	0	1	1	0	0	⋯	1	1	0.55
2	1	1	1	1	0	⋯	1	1	0.80
3	0	1	0	0	0	⋯	1	1	0.70
4	1	1	1	1	0	⋯	1	1	0.80
5	0	1	1	1	0	⋯	1	1	0.85
⋯	⋯	⋯	⋯	⋯	⋯	⋱	⋯	⋯	⋯
67	1	1	1	1	1	⋯	1	1	1.00
68	0	0	0	0	0	⋯	0	1	0.25
$\bar{X}_{Pi}$	0.71	0.91	0.74	0.72	0.31	⋯	0.91	0.96	0.74

Nota: Los datos completos se encuentran en el archivo suplementario Examen A. csv. Los símbolos $\bar{X}_{pI}$ y $\bar{X}_{Pi}$ representan la media de la fila (del sustentante *p*) y de la columna (de la pregunta *i*), respectivamente.

Examen B

Sustentantes (*p*)	Evaluadores (*j*) — Evaluador 1: Preguntas (*i*) 1	2	3	4	5	6	Evaluador 2: Preguntas (*i*) 1	2	3	4	5	6	$\bar{X}_{pIJ}$
1	7	8	7	6	8	7	6	7	6	7	7	8	7.0
2	10	9	6	7	7	7	7	7	8	8	8	8	7.7
3	9	9	8	8	8	8	7	8	7	9	6	8	7.9
4	9	10	9	8	8	10	9	7	9	10	9	10	8.9
5	9	9	9	8	8	8	8	7	8	9	9	9	8.4
⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
30	8	10	8	8	10	9	8	9	8	9	8	8	8.6
31	9	10	9	8	8	9	9	9	7	10	9	9	8.8
$\bar{X}_{Pij}$	8.5	6.9	7.6	7.3	7.5	8.1	8.9	7.1	8.2	7.2	8.1	8.4	7.8

Nota: Los datos completos se encuentran en el archivo suplementario Examen B.csv. Los símbolos $\bar{X}_{pIJ}$ y $\bar{X}_{Pij}$ representan la media de la fila (del sustentante *p*) y de la columna (de la pregunta *i* y el evaluador *j*), respectivamente.

El Examen B es una evaluación que aplica un profesor de la licenciatura en Filosofía a los 31 estudiantes de su grupo de la asignatura de Ética. Este examen consiste en seis preguntas que

requieren que los sustentantes reflexionen brevemente sobre ciertos temas (por ejemplo, con una crítica sobre cierta teoría o una comparación de diferentes perspectivas éticas). Tienen que responder a cada pregunta con un breve ensayo (de una página aproximadamente). Para asignar una calificación a las respuestas en cada pregunta, el profesor pide la ayuda a dos de sus estudiantes de doctorado (que trabajan con él como ayudantes de profesor); después de revisar junto con ellos los criterios para calificar las respuestas, les pide que ambos valoren, de manera independiente y utilizando una escala de 0 a 10, la respuesta en cada una de las seis preguntas de cada estudiante. (El profesor es consciente de que el juicio sobre la calidad de la reflexión que ha mostrado el sustentante en sus respuestas del examen hasta cierto punto es subjetivo; esta es la razón por la que pide una calificación independiente por sus ayudantes.) La calificación en el examen es el promedio en los 12 puntajes que se tiene para el sustentante (seis puntajes del primer evaluador más seis puntajes del segundo). Los datos resultantes se estructuran como se muestra en la parte inferior de la Tabla 1.

CONCEPTOS CENTRALES DE LA TEORÍA DE LA GENERALIZABILIDAD

La Teoría G consiste en un marco de referencia conceptual y una metodología que permite al investigador desentrañar múltiples fuentes de error y estimar su contribución relativa a las mediciones realizadas. Los orígenes de esta teoría se encuentran en la TCT y en el análisis de la varianza (ANOVA, por sus siglas en inglés).

En la Teoría G se describen las mediciones que se realizan a las personas en los términos de las condiciones bajo las cuales se han observado. Estas condiciones forman las *facetas*. En el Examen A, las mediciones son los puntajes de los sustentantes en las 20 preguntas del examen; estas 20 preguntas son las 20 *condiciones* de la *faceta* "preguntas" (la cual es la única faceta considerada para este examen). El Examen B, en el cual se observan los puntajes otorgados a las respuestas de los 31 individuos por cada uno de los dos evaluadores en las seis preguntas del examen, presenta un diseño con dos facetas: la faceta "preguntas" (con seis condiciones), y la faceta "evaluadores" (con dos condiciones). En ambos ejemplos, los sustentantes en el examen son el *objeto de la medición* (queremos conocer de cada sustentante su nivel en el constructo evaluado por el examen) y no se consideran una faceta; este término se reserva para referirse a una fuente de error, es decir, a un factor que causa variabilidad en las distintas mediciones realizadas a un sustentante. En este sentido, las preguntas del examen son una faceta, por ejemplo, porque difieren en dificultad y, así, son una causa de la variabilidad en los puntajes en las distintas preguntas. De igual manera, los dos evaluadores en el Examen B definen una faceta, ya que pueden diferir en su juicio sobre las respuestas de los sustentantes, lo cual lleva a diferencias en los puntajes observados.

No solo el número de facetas consideradas define el diseño, sino también la manera en que se combinan (entre sí y con las personas). En el Examen A, los 68 sustentantes responden a las mismas 20 preguntas, lo cual se conoce como un *diseño cruzado,* denotado como "personas×preguntas". Al contrario, un *diseño anidado* se presentaría cuando cada sustentante en el examen respondiera a un conjunto de preguntas únicas, es decir, distintas de las

preguntas de los demás. Un ejemplo de este tipo de diseño ocurre en los exámenes orales, donde los sustentantes realizan su examen por turnos y cada sustentante recibe, por ejemplo, tres preguntas, diferentes de las preguntas de los demás (para evitar que los últimos sustentantes del examen aprendan las preguntas de los que pasaron primero). Este diseño se denota como "reactivos : personas" ("reactivos anidados en personas"). En diseños con dos o más facetas, ciertas facetas pueden estar cruzadas y otras anidadas. En el Examen B, por ejemplo, tenemos un diseño de dos facetas totalmente cruzadas: para cada sustentante, sus respuestas en cada una de las seis preguntas son calificadas por cada uno de los dos ayudantes del profesor titular ("personas×preguntas×evaluadores"), llevando a 12 mediciones para cada sustentante. Si, al contrario, el profesor titular hubiera decidido que su primer ayudante evaluara las preguntas 1, 2 y 3 de cada sustentante y el segundo las preguntas 4, 5 y 6, entonces tendríamos solo seis observaciones por persona y un diseño con las preguntas anidadas en la faceta de evaluadores, que se denota como "personas × (preguntas : evaluadores)".

Los lectores familiarizados con ANOVA se habrán dado cuenta de las similitudes en la terminología que se utiliza para esta técnica y la que acabamos de definir para la Teoría G. Por ejemplo, también ANOVA distingue entre diseños cruzados y anidados; por otro lado, "factores" (o "variables independientes") en ANOVA son conceptualmente similares a las "facetas" en la Teoría G; y los niveles de cada factor en ANOVA corresponden a las condiciones de las facetas en la Teoría G. Veremos en las secciones que vienen que un análisis en el marco de la Teoría G implica un ANOVA a los datos observados. Cabe mencionar que, para comprender a cabalidad los análisis en las siguientes secciones, sirve tener una introducción a ANOVA, como se encuentra, por ejemplo, en los libros de Pardo y San Martín (2010) y Tejedor (2019).

En el caso del Examen A observamos, para cada sustentante, sus respuestas en las 20 preguntas que conforman el examen. No se duda de que el profesor podría haber utilizado otras 20 integrales como preguntas del examen. El conjunto de todas las posibles observaciones admisibles –en este ejemplo, los puntajes en todas las preguntas que el profesor considera aceptables para incluir en el examen– se llama en la Teoría G el *universo*. Idealmente, para determinar la habilidad matemática de un sustentante en el ejemplo del Examen A se observarían sus respuestas (y puntajes) en todas las preguntas del universo y se calcularía, en este universo de preguntas, su calificación (porcentaje de respuestas correctas). Esta calificación es la *puntuación universo* del sustentante. Es obvio que, por razones prácticas, solo se pueden observar las respuestas en una *muestra* de preguntas y utilizamos, para conocer la habilidad matemática de los sustentantes, su *puntuación observada* en esta muestra de preguntas. Esto significa que hacemos una generalización a partir de la puntuación observada para conocer la puntuación universo. La precisión con la que las puntuaciones observadas pueden generalizarse a la puntuación universo se llama la generalizabilidad y, como se ilustrará en las siguientes secciones, se cuantifica a través del *coeficiente de generalizabilidad*.

Si se aplican las ideas anteriores al Examen B, el universo se define igualmente como todas las observaciones admisibles para el grupo de sustentantes considerado. En este caso, el universo se define por dos facetas: no solo consideramos todas las posibles preguntas que

según el profesor titular son adecuadas para incluirse en el examen, sino también a todos los posibles evaluadores que podría haber involucrado para calificar las respuestas de los sustentantes. Las puntuaciones universo, en este caso, son las puntuaciones que se habrían obtenido si todos los evaluadores admisibles hubieran calificado las respuestas de los sustentantes en todas las preguntas admisibles para el examen. Como esto no es viable en la práctica, se hace una generalización a partir de la puntuación observada que se calcula con base en la muestra de 12 mediciones disponibles para cada sustentante.

Un análisis en el marco de la Teoría G típicamente consiste en dos estudios. En primera instancia, se lleva a cabo el *estudio de generalizabilidad* (Estudio G), el cual estima la contribución a la varianza en las mediciones observadas de (a) las habilidades evaluadas de los sustentantes, junto con (b) las distintas facetas (fuentes de error) consideradas. Posteriormente, se utilizan los resultados del Estudio G en el *estudio de decisión* (Estudio D) para diseñar de la mejor manera posible un examen con un propósito particular, o, de manera más general, un instrumento de medición. El Estudio D evalúa diferentes diseños para un examen a la luz de los objetivos, las limitaciones prácticas y las decisiones precisas que se quieren tomar a partir de los puntajes observados en el examen, minimizando de esta manera el error de medición y maximizando la confiabilidad del examen. En las siguientes secciones continuamos con los ejemplos guía (diseños con una faceta y dos facetas, respectivamente) e ilustramos cómo se llevan a cabo los Estudios G y D, así como los resultados que generan.

DISEÑO CON UNA FACETA

En esta sección elaboramos el ejemplo del Examen A, en el cual figura una faceta: la de las preguntas. Primero elaboramos el modelo teórico (explicando cómo se particiona la puntuación y la varianza observada en diferentes componentes) y posteriormente, con base en este modelo, realizamos e ilustramos los Estudios G y D para este ejemplo.

Descomposición de la Puntuación Observada

La medición que se realiza a la persona p a través de la pregunta i la representamos por X_{pi}. En la frase anterior, p e i se refieren, de una manera general, a cualquier persona de la población de personas que se consideren como posibles participantes al examen y a cualquier pregunta del universo de preguntas que sean admisibles para este examen, respectivamente. Para un diseño cruzado con una faceta el puntaje X_{pi} se descompone en cuatro componentes de la siguiente manera:

$$X_{pi} = \mu + \alpha_p + \beta_i + \varepsilon_{p\times i,e}, \tag{3}$$

$$\text{donde } \alpha_p \equiv \mu_p - \mu, \tag{3a}$$

$$\beta_i \equiv \mu_i - \mu, \tag{3b}$$

$$\varepsilon_{p\times i,e} \equiv X_{pi} - \mu_p - \mu_i + \mu. \tag{3c}$$

El primero de estos componentes, μ, representa a la media global de todas las mediciones que resultan de aplicar todas las preguntas del universo a todas las personas de la población. En la Ecuación 3, μ recoge todas las influencias constantes, es decir, el efecto de todos los factores que siempre tienen la misma influencia en la medición X_{pi}, independientemente de la persona o de las preguntas o de otros factores con un efecto variable (como el error de medición). Por ejemplo, considerando que el profesor es el único que califica las respuestas al Examen A y suponiendo que en cada medición (de cualquier persona y de cualquier pregunta) aplica los mismos criterios con la misma rigidez, el profesor es uno de los efectos que contribuyen al efecto global μ.

Con respecto al segundo componente, α_p, la Ecuación 3a lo define como la diferencia entre μ_p y μ, donde μ_p es el puntaje universo de la persona p; en otras palabras, μ_p es el promedio de los puntajes que obtendría esta persona si le administráramos todas las preguntas del universo. Por consiguiente, α_p, que representa el efecto de la persona p, indica en qué medida el puntaje universo de esta persona se encuentra arriba de la media global (si es positivo) o abajo (si es negativo). Nótese que, por las definiciones anteriores, el promedio de los efectos α_p de todas las personas de la población es igual a 0 (ya que los efectos positivos y negativos contrarrestan).

De manera análoga, se define el tercer componente, β_i, como la diferencia entre μ_i y μ, donde μ_i es el promedio de los puntajes en la pregunta i (o bien, en caso de preguntas dicotómicas como en el Examen A, el porcentaje de personas que la acertarían) si se administrara a todas las personas de la población. De esta manera, μ_i se puede interpretar como la dificultad (o, más bien, la facilidad) de la pregunta i: preguntas con la μ_i más alta son más fáciles; si μ_i es más baja, la pregunta es más difícil. El efecto de la pregunta i, β_i, indica que esta pregunta es más fácil (si es positivo) o más difícil (si es negativo) que la dificultad promedio de las preguntas. En promedio, los efectos β_i de todas las preguntas en el universo es igual a 0.

Por último, el cuarto componente en la Ecuación 3, $\varepsilon_{p\times i,e}$, recoge el efecto residual; es decir, el efecto de todos los factores que influyen en la medición X_{pi}, excepto el efecto global μ, el de la persona α_p, y el de la pregunta β_i. Además del efecto del error de medición (es decir, los factores con una influencia no sistemática que no se contemplaron explícitamente en el diseño), el efecto residual incluye también el efecto de la interacción entre la persona y la pregunta. Es la razón por la que, para este diseño cruzado con una faceta, se incluyen como subíndices de ε, tanto $p \times i$ (la interacción) como e (el error de medición). De la definición en la Ecuación 3c sigue que, para cualquier pregunta i, el promedio de los efectos residuales $\varepsilon_{p\times i,e}$ entre todas las personas de la población es 0, así como que el promedio de los $\varepsilon_{p\times i,e}$ entre todas las preguntas del universo es 0 para cualquier persona p.

Descomposición de la Varianza Observada

Como acabamos de explicar, el primer componente en la Ecuación 3 es constante, igual para todas las mediciones en el universo de preguntas y la población de personas. Por otro lado, el componente α_p varía: tiene un valor distinto para diferentes personas. Esto quiere decir que podemos definir la varianza σ_α^2, la cual representa la variabilidad con respecto al com-

ponente α_p en la población de personas. Una interpretación alternativa de σ_α^2 es que corresponde con la variabilidad de los puntajes universo entre las personas de la población. De forma similar, el componente β_i varía entre las distintas preguntas del universo y esta variabilidad se cuantifica en la varianza σ_β^2. Se puede interpretar esta varianza como la variabilidad en la dificultad entre todas las preguntas del universo. Finalmente, se define la varianza σ_ε^2, que representa la variabilidad residual (por la variabilidad en el error de medición y el efecto de interacción, considerando las distintas combinaciones de persona-pregunta en la población de personas y el universo de preguntas). Se puede derivar del modelo introducido en la Ecuación 3 que

$$\sigma_X^2 = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\varepsilon^2. \tag{4}$$

En palabras, la varianza en la colección de todas las mediciones (de todas las personas de la población a quienes se administraron todas las preguntas del universo) se particiona en tres: varianza que se debe a las personas (porque difieren en su nivel general, representado por su puntaje universo), varianza debida a las preguntas (porque difieren en su grado de dificultad) y varianza debida a otros factores (incluyendo la interacción entre personas y preguntas). Las tres varianzas se llaman componentes de varianza en la colección de mediciones X_{pi} consideradas en este diseño.

Estudio de Generalizabilidad

Debe ser claro que las Ecuaciones 3 y 4 son parte de un modelo teórico y que los componentes de la puntuación y varianza observada (μ, μ_p, α_p, σ_α^2, etc.) son cantidades teóricas que en la práctica son desconocidas. Con base en los datos observados, como los que se muestran en la Tabla 1, podemos obtener estimaciones de estos componentes. En particular, el Estudio G se centra en la estimación de los componentes de varianza de la Ecuación 4.

El primer paso del Estudio G consiste en realizar un ANOVA con base en el modelo correspondiente al diseño del estudio (que se describió en las subsecciones anteriores). La Tabla 2 resume los resultados principales del ANOVA realizado a los datos del Examen A que se necesitan para estimar los componentes de varianza. En particular, las primeras cuatro columnas presentan los resultados que cualquier programa de análisis estadístico generaría para el ANOVA a estos datos. (El archivo suplementario Script_Análisis.R realiza los análisis en el entorno del software R). La última columna contiene las estimaciones de los componentes de varianza, las cuales se derivan a partir de las medias cuadráticas (MS, por sus siglas en inglés) mostradas en la penúltima columna, aplicando las siguientes fórmulas:

$$\hat{\sigma}_{\alpha}^{2} = \frac{MS_{\alpha} - MS_{\varepsilon}}{n_i} = \frac{1.0142 - 0.1259}{20} = 0.0444 \quad (5a)$$

$$\hat{\sigma}_{\beta}^{2} = \frac{MS_{\beta} - MS_{\varepsilon}}{n_p} = \frac{2.2373 - 0.1259}{68} = 0.0310 \quad (5b)$$

$$\hat{\sigma}_{\varepsilon}^{2} = MS_{\varepsilon} = 0.1259 \quad (5c)$$

En estas fórmulas, n_i y n_p son el número de preguntas y el número de personas, respectivamente, en los datos que se utilizan para la estimación. El símbolo ^ encima de los componentes de varianza indica que se trata de estimaciones (que se distinguen de los valores teóricos de la Ecuación 4).

Puesto que (las estimaciones para) los componentes de varianza dependen del esquema de calificación que se utilizó para las preguntas (en este caso, puntajes dicotómicos, 0 y 1), los valores obtenidos son difíciles de interpretar. Es la razón por la que se calcula el porcentaje de varianza que cada componente contribuye a la varianza (total) observada σ_X^2. (Estos porcentajes se muestran también en la última columna de la Tabla 2.) Por ejemplo, para la varianza σ_α^2, la cual es la varianza de los puntajes universo de las personas, se estima que contribuye 0.0444 / (0.0444 + 0.0310 + 0.1259) ≈ .221 (es decir, el 22.1%) a la varianza total en las mediciones (los puntajes 0 y 1, considerando el universo de preguntas y la población de personas). Los resultados del Estudio G enseñan que la contribución de la varianza residual (62.5%) es relativamente grande en comparación con la contribución de la varianza entre personas (22.1%) y entre preguntas (15.4%).

Estudio de Decisión

Hasta este momento, el modelo y análisis del Examen A se ha enfocado en la descomposición de la medición X_{pi}; es decir, el puntaje de la persona en una pregunta individual del universo. Sin embargo, el profesor no toma decisiones sobre sus estudiantes con base en sus puntajes en una pregunta, sino en sus promedios (o porcentajes de respuestas correctas) de todas las preguntas que conforman el examen. Este promedio de una persona p en las 20 preguntas del examen la denotamos como $\bar{X}_{pI}$ (con *i* mayúscula para aclarar que se trata no de una pregunta, sino de un conjunto de preguntas). La última columna de la Tabla 1 muestra los puntajes $\bar{X}_{pI}$ de los sustentantes.

El profesor podría haber elegido otras 20 preguntas del universo para el examen; en este caso se tendría un *examen aleatoriamente paralelo*; seguramente, las mediciones $\bar{X}_{pI}$ en este examen paralelo hubieran sido diferentes. Incluso, en vez de haber construido un examen de 20 preguntas, el profesor podría haber optado por un examen de un número diferente de preguntas; de manera general, el Estudio D responde a la siguiente pregunta sobre los

exámenes de n_i preguntas,[1] que se seleccionen de manera aleatoria del universo de preguntas: ¿con qué precisión permite la puntuación observada $\bar{X}_{pI}$ (el promedio en las n_i preguntas del examen) generalizarse al universo de todos los exámenes aleatoriamente paralelos de n_i preguntas? Este universo se llama el *universo de generalización* del Estudio D.

En el Estudio D, se consideran entonces las puntuaciones $\bar{X}_{pI}$ en exámenes paralelos de n_i preguntas en vez de los puntajes X_{pi} en preguntas individuales; con respecto a la varianza de estas puntuaciones $\bar{X}_{pI}$ (la variabilidad en las puntuaciones promedio en el universo de exámenes paralelos y en la población de personas) se puede derivar que:

$$\sigma_{\bar{X}}^2 = \sigma_\alpha^2 + \frac{\sigma_\beta^2}{n_i} + \frac{\sigma_\varepsilon^2}{n_i}, \tag{6}$$

donde las varianzas σ_α^2, σ_β^2, σ_ε^2 se definen de manera idéntica que anteriormente en la Ecuación 4. El término σ_β^2/n_i indica la varianza con respecto a la dificultad (el puntaje promedio en las n_i preguntas de todas las personas en la población) entre todos los distintos exámenes aleatoriamente paralelos en el universo.[2] De la misma manera, σ_ε^2/n_i se refiere a la varianza residual, otra vez, considerando a todos los exámenes aleatoriamente paralelos de n_i preguntas en la población de personas. Nótese que, al comparar las Ecuaciones 4 y 6, las puntuaciones universo de las personas (σ_α^2) tienen una contribución relativamente más grande a la varianza de los puntajes promedio en los exámenes (dado que la varianza de la dificultad y del efecto residual es más pequeña en el caso de conjuntos de preguntas en vez de preguntas individuales).

Para derivar el coeficiente de generalizabilidad, es importante precisar el objetivo de las decisiones que se desean tomar con base en las puntuaciones $\bar{X}_{pI}$ de los distintos sustentantes en el examen. Se distingue entre dos tipos de objetivos: primero, el profesor podría utilizar los puntajes en el examen para situar el desempeño de cada sustentante *relativo* al desempeño de los otros sustentantes. Un ejemplo de esta situación se presentaría si el profesor, con base en los puntajes del examen, quisiera seleccionar a los ocho mejores estudiantes para posteriormente invitarles a participar en la primera etapa de las Olimpiadas Mexicanas de Matemáticas (ya que solo un máximo de ocho personas de su grupo puede inscribirse). El objetivo principal en este caso es construir un *ranking* de los sustentantes en el examen y se utilizarán las puntuaciones en el Examen A para decisiones relativas. El

---

1 Lo anterior quiere decir que, a pesar de que el número de preguntas que el profesor incluyó en el Examen A es 20, el Estudio D permite contemplar qué pasaría (o hubiera pasado) si el examen consistiera en un número distinto de preguntas. Entonces, n_i para el Estudio D puede ser diferente de 20. Esto contrasta con el Estudio G, donde n_i siempre es el número de preguntas en los datos que se utilizan para el ANOVA.

2 Los lectores familiarizados con la estadística inferencial habrán notado la analogía con el resultado bien conocido que la varianza de la media muestral es igual a la varianza de la variable original entre n, el tamaño de la muestra: $\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}$.

coeficiente de generalizabilidad para decisiones relativas para el diseño actual se estima a través de:

$$\hat{\rho}^2_{\text{rel}} = \frac{\hat{\sigma}^2_{\alpha}}{\hat{\sigma}^2_{\alpha} + \frac{\hat{\sigma}^2_{\varepsilon}}{n_i}}. \tag{7}$$

Introduciendo en esta fórmula los resultados del Estudio G (véase la Tabla 2), se obtiene directamente la siguiente estimación del coeficiente de generalizabilidad para el examen de 20 preguntas que el profesor aplicó:

$$\hat{\rho}^2_{\text{rel}} = \frac{0.0444}{0.0444 + \frac{0.1259}{20}} = .876.$$

Tabla 2. Tabla con los resultados del análisis de varianza y el estudio de generalizabilidad aplicados a los datos del Examen A

Efectos	Sumas cuadráticas	Grados de libertad	Medias cuadráticas	Estimaciones de los componentes de varianza
Personas (α_p)	67.949	67	1.0142	$\hat{\sigma}^2_{\alpha} = 0.0444$ (22.1%)
Preguntas (β_i)	42.508	19	2.2373	$\hat{\sigma}^2_{\beta} = 0.0310$ (15.4%)
Residual ($\varepsilon_{p\times i,e}$)	160.242	1273	0.1259	$\hat{\sigma}^2_{\varepsilon} = 0.1259$ (62.5%)

Nota. El archivo suplementario Script_Análisis. R contiene el código que permite llevar a cabo los cálculos en el entorno del *software* R para los resultados mostrados.

En cuanto a la interpretación de este coeficiente, cabe mencionar que valores más cercanos a 1 indican que las diferencias observadas entre los distintos participantes en el examen reflejan relativamente más diferencias entre sus puntajes universo. Para el resultado obtenido para el Examen A, casi el 88% de las diferencias observadas en el examen se deben a diferencias entre las personas (el resto, se atribuye al error de medición y/o la interacción particular entre personas y las preguntas del examen).

De manera alternativa, el objetivo del profesor podría consistir en determinar para cada persona si su desempeño alcanza cierto nivel mínimo, definido a partir de cierto criterio o norma; más concreto, el profesor puede estar interesado en saber de cada estudiante si tiene el bagaje suficiente para pasar al curso avanzado de Cálculo Diferencial e Integral del siguiente semestre e identificar a aquellos estudiantes que necesitan ejercicios adicionales

para pasar al siguiente semestre. Por ejemplo, podría ser que los estudiantes que dominaran menos del 75% de las preguntas en el universo (es decir, estudiantes con un puntaje universo menor de 75%), según la opinión del profesor, necesitan hacer una tarea adicional. En este caso, el profesor no está interesado en el desempeño de cada estudiante en comparación con otros sustentantes, sino en su nivel absoluto de desempeño; en el marco de la Teoría G, se dice que las puntuaciones en el examen se utilizan para tomar decisiones absolutas. El *coeficiente de generalizabilidad para decisiones absolutas* para el diseño actual se estima por:

$$\hat{\rho}^2_{\text{abs}} = \frac{\hat{\sigma}^2_\alpha}{\hat{\sigma}^2_\alpha + \frac{\hat{\sigma}^2_\beta}{n_i} + \frac{\hat{\sigma}^2_\varepsilon}{n_i}}, \tag{8}$$

Io cual, introduciendo los resultados del Estudio G, produce la siguiente estimación para el Examen A:

$$\hat{\rho}^2_{\text{abs}} = \frac{0.0444}{0.0444 + \frac{0.0310}{20} + \frac{0.1259}{20}} = .850.$$

Para este coeficiente, valores más cercanos a 1 indican que las diferencias entre el puntaje observado en el examen de los respectivos sustentantes y un criterio absoluto reflejan más diferencias entre la puntuación universo de los sustentantes y este criterio. En el caso del Examen A, 85% de estas diferencias entre el puntaje (porcentaje de respuestas correctas) en el examen y el criterio de 75% para decidir que el estudiante debe hacer una tarea adicional reflejan diferencias entre su puntaje universo y este mismo criterio.

Concluimos esta sección con dos comentarios. Primero, uno puede preguntarse cuál es la lógica de tener fórmulas distintas para decisiones relativas y absolutas. Comparando las Ecuaciones 7 y 8, se nota que la única diferencia es que la primera no incluye en el denominador la varianza del grado de dificultad entre los distintos exámenes aleatoriamente paralelos (σ^2_β / n_i). La razón es que, si solo nos interesa saber qué tanto Juan es mejor que Pedro, entonces no importa si el examen (que ambos contestaron) es un examen fácil o difícil. Si el examen fuera fácil, ambos tendrían un puntaje más alto en comparación con su puntaje en un examen difícil, pero la Teoría G espera que la *diferencia* entre ambos sea la misma en ambos casos, independientemente de la dificultad del examen. Retomando el ejemplo anterior: se espera que los ocho estudiantes con mejores puntajes serían los mismos, tanto si el examen fuera fácil como si fuera difícil. Por otro lado, si el interés está en una decisión absoluta, sí es importante la dificultad del examen: obviamente, si el examen resultara más difícil, los sustentantes no alcanzarían de igual manera el criterio de 75% de respuestas correctas. Este razonamiento explica la diferencia entre los coeficientes de generalizabilidad para decisiones relativas y absolutas.

Segundo, las Ecuaciones 7 y 8 muestran de una manera muy clara el efecto de aumentar (o disminuir) el número de preguntas (n_i) en el examen. El lector puede verificar que, si el Examen A consistiera en 50 en vez de 20 preguntas, los coeficientes de generalización aumentarían. En este sentido, las fórmulas de la Teoría G aquí presentadas se relacionan con la famosa fórmula de Spearman-Brown que expresa la relación entre la longitud de una prueba y la confiabilidad en la TCT.

DISEÑO CON DOS FACETAS

En esta sección, describimos el análisis para un diseño cruzado con dos facetas, elaborando el ejemplo del Examen B. En esta elaboración, nos enfocaremos en estos elementos para los cuales el diseño con dos facetas difiere o añade algo más en comparación con el diseño con una faceta. Mantenemos el mismo orden para el desarrollo que en la sección anterior: empezamos con el modelo en que se basan los análisis, incluyendo la descomposición de la puntuación observada y la varianza observada, y posteriormente delineamos los pasos del Estudio G y el Estudio D.

Descomposición de la Puntuación y la Varianza Observada

En el caso del Examen B consideramos, de manera genérica, el puntaje X_{pij} de la persona p en la pregunta i otorgado por el evaluador j (donde, como comentamos anteriormente, las preguntas y los evaluadores que se reconocen como adecuados para este examen definen el universo de mediciones admisibles para la población de personas). En el diseño cruzado de dos facetas, X_{pij} se expresa como sigue:

$$X_{pij} = \mu + \alpha_p + \beta_i + \gamma_j + (\alpha\beta)_{pi} + (\alpha\gamma)_{pj} + (\beta\gamma)_{ij} + \varepsilon_{p\times i\times j,e}, \qquad (9)$$

donde $\alpha_p \equiv \mu_p - \mu,$

$$\beta_i \equiv \mu_i - \mu,$$

$$\gamma_j \equiv \mu_j - \mu,$$

$$(\alpha\beta)_{pi} \equiv \mu_{pi} - \mu_p - \mu_i + \mu,$$

$$(\alpha\gamma)_{pj} \equiv \mu_{pj} - \mu_p - \mu_j + \mu,$$

$$(\beta\gamma)_{ij} \equiv \mu_{ij} - \mu_i - \mu_j + \mu,$$

$$\varepsilon_{p\times i\times j,e} \equiv X_{pij} - \mu_{pi} - \mu_{pj} - \mu_{ij} + \mu_p + \mu_i + \mu_j - \mu.$$

La media global μ se define como el promedio de la medición entre todas las personas de la población y todas las preguntas y todos los evaluadores del universo. Por otro lado, μ_p es el puntaje universo de la persona *p*; es decir, el promedio de sus puntajes en todas las preguntas, evaluadas por todos los evaluadores. Similar al diseño con una faceta, α_p es el efecto de la persona *p* e indica cómo el puntaje universo de la persona *p* se relaciona con la media global. De manera análoga, se definen μ_i, μ_j, β_i y γ_j: μ_i es la dificultad de la pregunta *i* (el promedio de los puntajes en esta pregunta de todas las personas de la población, considerando los juicios de todos los evaluadores del universo, así que valores más altos en μ_i indican una pregunta más fácil) y μ_j la exigencia del evaluador *j* (que es el promedio de todos sus puntajes otorgados a las respuestas de todas las personas de la población en todas las preguntas del universo, de tal manera que los evaluadores con μ_j más alta son menos exigentes). Las β_i y γ_j, similar a α_p, indican la diferencia de la pregunta *i* y el evaluador *j*, respectivamente, con la media global μ (la cual se puede interpretar como la dificultad promedio de todas las preguntas, así como la exigencia promedio de todos los evaluadores en el universo).

Como diferencia importante con el diseño descrito en la sección anterior, el diseño cruzado con dos facetas incluye varios términos de interacción (separados del efecto residual). En primera instancia, la interacción $(\alpha\beta)_{pi}$ da cuenta del efecto de la *combinación particular* de la persona *p* y la pregunta *i*. Por ejemplo, puede ser que, a pesar de que la pregunta *i* es difícil (con un efecto β_i negativo) y que la persona *p* tiene un nivel debajo de la media de la población (con un efecto α_p negativo), la pregunta *i* en particular le resulta más fácil a la persona *p* (por las experiencias previas de la persona). Las interacciones $(\alpha\gamma)_{pj}$ y $(\beta\gamma)_{ij}$ se interpretan de una manera similar: $(\alpha\gamma)_{pj}$ alude a la posibilidad que los evaluadores pueden ser más exigentes con unas personas que con otras. (Así, por ejemplo, en caso de favoritismos de ciertos evaluadores a ciertos sustentantes, se espera un efecto de interacción $(\alpha\gamma)_{pj}$ importante.) El efecto de interacción $(\beta\gamma)_{ij}$ significa que un evaluador califica las respuestas en ciertas preguntas de manera más o menos exigente que aquellas en otras preguntas (por ejemplo, porque está más o menos familiarizado con el tema elaborado en la pregunta o porque no entendió cabalmente los criterios de evaluación para juzgar las respuestas). La definición de estos efectos de interacción incluye a las medias μ_{pi} (el promedio de los puntajes otorgados por todos los evaluadores en el universo a la respuesta en la pregunta *i* dada por la persona *p*), μ_{pj} (el promedio de los puntajes entre todas las preguntas del universo, respondidas por la persona *p* y valoradas por el evaluador *j*) y μ_{ij} (el promedio de los puntajes de todas las personas en la pregunta *i* valoradas por el evaluador *j*). Por último, el efecto residual en el diseño cruzado con dos facetas incluye el efecto de la interacción triple entre personas, preguntas y evaluadores (que es difícil de interpretar y tiene menos importancia práctica), mezclado con el error de medición.

Como siguiente paso, se considera la variabilidad entre todas las mediciones X_{pij} (es decir, de todas las personas en la población y de todas las preguntas y todos los evaluadores en el universo), expresada por la varianza σ_X^2, la cual con base en las definiciones anteriores se particiona en siete componentes de varianza como sigue:

$$\sigma_X^2 = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_\gamma^2 + \sigma_{\alpha\beta}^2 + \sigma_{\alpha\gamma}^2 + \sigma_{\beta\gamma}^2 + \sigma_\varepsilon^2. \tag{10}$$

En palabras, la varianza total de las mediciones se explica por diferencias entre las personas con respecto a su puntaje universo (σ_α^2), diferencias entre preguntas con respecto a su grado de dificultad (σ_β^2), diferencias entre evaluadores con respecto a su nivel de exigencia (σ_γ^2), diferencias entre las combinaciones persona-pregunta ($\sigma_{\alpha\beta}^2$), diferencias entre las combinaciones persona-evaluador ($\sigma_{\alpha\gamma}^2$), diferencias entre las combinaciones pregunta-evaluador ($\sigma_{\beta\gamma}^2$) y diferencias residuales (σ_ε^2). En la siguiente sección explicamos cómo se estiman estos componentes de varianza a partir de datos observados.

Estudio de Generalizabilidad

Para el Estudio G realizamos un ANOVA, con base en el modelo de la Ecuación 9, a los datos empíricos para el Examen B que se muestran en la parte inferior de la Tabla 1. El resultado de este análisis se encuentra en las primeras cuatro columnas de la Tabla 3. La última columna contiene las estimaciones de los componentes de varianza, que en este caso se calcularon a partir de las medias cuadráticas de la siguiente manera:

$$\hat{\sigma}_\alpha^2 = \frac{MS_\alpha - MS_{\alpha\beta} - MS_{\alpha\gamma} + MS_\varepsilon}{n_i \cdot n_j} = \frac{6.0443 - 0.8036 - 0.5023 + 0.4505}{6 \cdot 2} = 0.4324$$

$$\hat{\sigma}_\beta^2 = \frac{MS_\beta - MS_{\alpha\beta} - MS_{\beta\gamma} + MS_\varepsilon}{n_p \cdot n_j} = \frac{23.9914 - 0.8036 - 1.1849 + 0.4505}{31 \cdot 2} = 0.3622$$

$$\hat{\sigma}_\gamma^2 = \frac{MS_\gamma - MS_{\alpha\gamma} - MS_{\beta\gamma} + MS_\varepsilon}{n_p \cdot n_i} = \frac{8.4301 - 0.5023 - 1.1849 + 0.4505}{31 \cdot 6} = 0.0387$$

$$\hat{\sigma}_{\alpha\beta}^2 = \frac{MS_{\alpha\beta} - MS_\varepsilon}{n_j} = \frac{0.8036 - 0.4505}{2} = 0.1766$$

$$\hat{\sigma}_{\alpha\gamma}^2 = \frac{MS_{\alpha\gamma} - MS_\varepsilon}{n_i} = \frac{0.5023 - 0.4505}{6} = 0.0086$$

$$\hat{\sigma}_{\beta\gamma}^2 = \frac{MS_{\beta\gamma} - MS_\varepsilon}{n_p} = \frac{1.1849 - 0.4505}{31} = 0.0237$$

$$\hat{\sigma}_\varepsilon^2 = MS_\varepsilon = 0.4505$$

En la Tabla 3 se muestran también las contribuciones relativas (como porcentajes) de cada componente de varianza a la varianza total. Se estima que el 29% de la varianza total en las mediciones se debe a diferencias entre (los puntajes universo de) las personas. También las diferencias entre (las dificultades de) las preguntas explican una parte importante (24.3%) en la varianza total. Además, el componente ($\sigma^2_{\alpha\beta}$) tiene una contribución relativamente grande (11.8%) a la varianza total, lo cual apunta a una interacción importante entre las personas y las preguntas, es decir, que las personas difieren con respecto a cuáles preguntas les resultan más difíciles. Por otro lado, las diferencias entre los evaluadores en cuanto a su exigencia son relativamente pequeñas (solo explican el 2.6% de la varianza total) y también las interacciones de los evaluadores con las personas ($\sigma^2_{\alpha\gamma}$) y con las preguntas ($\sigma^2_{\beta\gamma}$) son diminutas (0.6% y 1.6%, respectivamente), por lo que se concluye que la exigencia de los evaluadores no varía mucho entre diferentes personas o para diferentes preguntas.

Tabla 3. Resultados del análisis de varianza y el estudio de generalizabilidad aplicados a los datos del Examen B

Efectos	Sumas cuadráticas	Grados de libertad	Medias cuadráticas	Estimaciones de los componentes de varianza	
Personas [α_p]	181.33	30	6.0443	$\hat{\sigma}^2_{\alpha} = 0.4324$	(29.0%)
Preguntas [β_i]	119.96	5	23.9914	$\hat{\sigma}^2_{\beta} = 0.3622$	(24.3%)
Evaluadores [γ_j]	8.43	1	8.4301	$\hat{\sigma}^2_{\gamma} = 0.0387$	(2.6%)
Personas × Preguntas [$(\alpha\beta)_{pi}$]	120.54	150	0.8036	$\hat{\sigma}^2_{\alpha\beta} = 0.1766$	(11.8%)
Personas × Evaluadores [$(\alpha\gamma)_{pj}$]	15.07	30	0.5023	$\hat{\sigma}^2_{\alpha\gamma} = 0.0086$	(0.6%)
Preguntas × Evaluadores [$(\beta\gamma)_{ij}$]	5.92	5	1.1849	$\hat{\sigma}^2_{\beta\gamma} = 0.0237$	(1.6%)
Residual ($\varepsilon_{p\times i\times j,e}$)	67.58	150	0.4505	$\hat{\sigma}^2_{\varepsilon} = 0.4505$	(30.2%)

Nota. El archivo suplementario Script_Análisis.R contiene el código que permite llevar a cabo los cálculos en el entorno del *software* R para los resultados mostrados.

Estudio de Decisión

Al tomar decisiones sobre los sustentantes a partir de sus puntajes en el Examen B, es oportuno tener en cuenta que habrían sido igualmente apropiados otras seis preguntas y otros dos evaluadores del universo de preguntas y evaluadores admisibles. Esto quiere decir que el universo de generalización del Estudio D para este caso es el universo de todos los exámenes aleatoriamente paralelos de seis preguntas y dos evaluadores.

Si es nuestro interés tomar decisiones relativas con respecto a los sustentantes, el coeficiente de generalizabilidad para este diseño cruzado de dos facetas se obtiene por:

$$\hat{\rho}^2_{\text{rel}} = \frac{\hat{\sigma}^2_{\alpha}}{\hat{\sigma}^2_{\alpha} + \frac{\hat{\sigma}^2_{\alpha\beta}}{n_i} + \frac{\hat{\sigma}^2_{\alpha\gamma}}{n_j} + \frac{\hat{\sigma}^2_{\varepsilon}}{n_i \cdot n_j}}, \qquad (11)$$

lo cual con las estimaciones de los componentes de varianza sacadas del Estudio G resulta en:

$$\hat{\rho}^2_{\text{rel}} = \frac{0.4324}{0.4324 + \frac{0.1766}{6} + \frac{0.0086}{2} + \frac{0.4505}{6 \cdot 2}} = .858.$$

Nótese que en la Ecuación 11, el denominador no incluye la varianza en la dificultad de las preguntas (σ^2_{β}), ni la varianza en la exigencia entre evaluadores (σ^2_{γ}), ni la varianza debida a la interacción entre estas dos facetas $(\sigma^2_{\beta\gamma})$. Efectivamente, la dificultad global de las preguntas o la exigencia general de los evaluadores no importa al tomar decisiones relativas: si el examen paralelo consistiera, por ejemplo, en preguntas más difíciles y/o las respuestas fueran valoradas por evaluadores más exigentes, el *orden* relativo entre los sustentantes no cambiaría (aunque todos los sustentantes tendrían puntajes más bajos). Por otro lado, las interacciones entre personas y preguntas $(\sigma^2_{\alpha\beta})$ y personas y evaluadores $(\sigma^2_{\alpha\gamma})$ sí entran como varianza error para la toma de decisiones relativas: si resulta que cierta pregunta es difícil para una persona, pero fácil para otra, entonces sí las preguntas incluidas en el examen afectarían el orden entre los sustentantes, igual que la selección de evaluadores del universo afectaría este orden en el caso de que ciertos evaluadores fueran menos exigentes para unos sustentantes que para otros.

El coeficiente de generalizabilidad para decisiones absolutas en este diseño se calcula como:

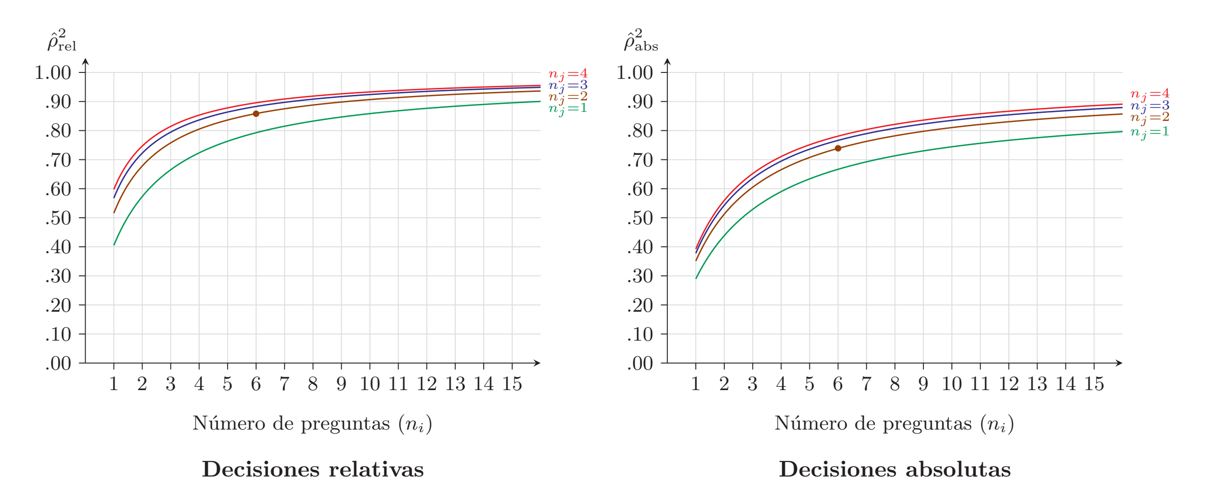
$$\hat{\rho}^2_{\text{abs}} = \frac{\hat{\sigma}^2_{\alpha}}{\hat{\sigma}^2_{\alpha} + \frac{\hat{\sigma}^2_{\beta}}{n_i} + \frac{\hat{\sigma}^2_{\gamma}}{n_j} + \frac{\hat{\sigma}^2_{\alpha\beta}}{n_i} + \frac{\hat{\sigma}^2_{\alpha\gamma}}{n_j} + \frac{\hat{\sigma}^2_{\beta\gamma}}{n_i \cdot n_j} + \frac{\hat{\sigma}^2_{\varepsilon}}{n_i \cdot n_j}} \qquad (12)$$

y para el ejemplo del Examen B es igual a:

$$\hat{\rho}^2_{\text{abs}} = \frac{0.4324}{0.4324 + \frac{0.3622}{6} + \frac{0.0387}{2} + \frac{0.1766}{6} + \frac{0.0086}{2} + \frac{0.0237}{6 \cdot 2} + \frac{0.4505}{6 \cdot 2}} = .739.$$

Para las decisiones absolutas, cuando el objetivo es saber si el sustentante alcanza cierto nivel de dominio, la dificultad de las preguntas y la exigencia de los evaluadores seleccionados sí afectan las decisiones: en exámenes con preguntas más difíciles y evaluadores más exigentes, los sustentantes requieren un nivel más alto para "aprobar" (es decir, para alcanzar el nivel mínimo de dominancia). En este caso, se observa una diferencia más grande (en comparación con el ejemplo anterior del diseño con una faceta) entre los coeficientes para decisiones relativas y absolutas, lo cual principalmente se debe a la variabilidad relativamente grande de las preguntas con respecto a su dificultad (σ_β^2) y el número relativamente bajo de preguntas en el examen. El efecto del número de preguntas (n_i) tanto como del número de evaluadores (n_j) en los coeficientes de generalizabilidad para este ejemplo se muestran en la Figura 1. Esta información permitirá al profesor titular de la materia tomar decisiones sobre la organización de su examen en el futuro, por ejemplo, si conviene más aumentar el número de preguntas o bien el número de evaluadores. Es claro que el número de preguntas tiene un efecto más importante en la generalizabilidad de los puntajes derivados de su examen. Por ejemplo, si el profesor quisiera tener un examen con un índice de generalizabilidad de 0.80 para tomar decisiones absolutas, se puede leer en la Figura 1 que le serviría aumentar el número de preguntas en el examen de seis a nueve.

Figura 1. Coeficientes de generalizabilidad para decisiones relativas (panel izquierdo) y absolutas (panel derecho) en función del número de preguntas (n_i) y el número de evaluadores (n_j) para el Examen B



OTROS ASPECTOS DE LOS DISEÑOS EN LA TEORÍA DE LA GENERALIZABILIDAD

En las secciones anteriores ilustramos los análisis para dos diseños típicos y relativamente sencillos en el marco de la Teoría G. La diferencia entre ambos diseños es fundamentalmente el número de facetas consideradas; obviamente, diseños multifacéticos generalmente llevan a análisis más complejos, pero conducen a resultados e interpretaciones más ricas. A continuación, examinamos brevemente otros dos aspectos que son relevantes al diseñar estudios de generalizabilidad.

En la sección de *Aspectos Centrales de la Teoría de la Generalizabilidad* ya explicamos la diferencia entre facetas cruzadas y facetas anidadas. Cruzar todas las condiciones de las facetas en un diseño multifacético muchas veces lleva a un número demasiado alto de combinaciones que, por las limitaciones logísticas, prácticas o éticas, no se pueden llevar a cabo. Por ejemplo, un estudio en el marco de la Teoría G del examen clínico objetivo estructurado (ECOE) que es común en las Ciencias de la Salud (véase, por ejemplo, Dizon et al., 2021; Espinosa-Vázquez et al., 2017; Trejo-Mejía et al., 2016) podría fácilmente considerar las siguientes facetas: estaciones (escenarios clínicos donde el sustentante se desempeña), examinadores (que evalúan el desempeño del sustentante), sitios (por ejemplo, clínicas) y versiones del examen (que se aplican en diferentes turnos). Es obviamente imposible que todos los evaluadores valoren el desempeño de todos los sustentantes en todas las estaciones de las distintas versiones del examen realizadas en las distintas clínicas. Más bien, los examinadores suelen estar anidados en las estaciones (cada estación tiene múltiples examinadores que solo evalúan esta estación) tanto como en los sitios (en cada clínica trabaja un equipo diferente de examinadores), mientras que las estaciones están anidadas en las versiones del examen (cada versión consiste en determinadas estaciones) pero se cruzan con los sitios (en cada sitio se lleva a cabo cada estación); por otro lado, los sustentantes están anidados en versiones y sitios (ya que cada estudiante participa solo en un turno y en un sitio). Precisamente porque múltiples fuentes de error pueden afectar el resultado de los sustentantes en los ECOEs, la Teoría G se ha convertido en el estándar de oro para estudiar la calidad psicométrica de estos exámenes.

Al desarrollar los ejemplos anteriores hemos considerado las condiciones de las facetas en los datos observados como una muestra aleatoria del universo de todas las preguntas y/o de todos los evaluadores admisibles. En lenguaje técnico se dice que se trata de *facetas aleatorias*. Aunque en la gran mayoría de las aplicaciones será apropiado considerar las facetas como aleatorias, hay ocasiones donde el objetivo del estudio no es generalizar las condiciones de cierta faceta a un universo más grande (o, más bien, que se considera que las condiciones del estudio conforman el universo completo). Un ejemplo se presentaría al incluir el turno (con las condiciones matutino y vespertino) como faceta en el estudio; probablemente el objetivo no sería generalizar a un universo que incluya otros turnos. En este caso, la faceta se considera *fija*. Cabe mencionar que la manera en que las facetas entran en el diseño (anidadas vs. cruzadas, aleatorias vs. fijas) tiene repercusiones en los análisis del Estudio G tanto como los del Estudio D. El lector interesado puede consultar las publicaciones de Brennan (2001) y Webb et al. (2006).

Aunque en los ejemplos anteriores se manejó el mismo diseño (con una o dos facetas cruzadas) tanto para el Estudio G como el Estudio D, es posible considerar distintos diseños para ambos estudios. Sin embargo, aplican restricciones; en general, son preferibles los diseños cruzados para los datos recopilados que se analizan ya que estos permiten, en el Estudio G, estimar todos los posibles componentes de varianza, mientras que, en el Estudio D, se puede evaluar cómo la anidación de una faceta en otra afecta los coeficientes de generalizabilidad. Por el contrario, no es posible estimar, a partir de datos recopilados

según un diseño anidado, los resultados que se obtuviesen en un diseño cruzado. Webb et al. (2006) describen cuáles diseños se pueden analizar en el Estudio D para cada uno en una lista de diseños para el ANOVA en el Estudio G.

COMENTARIOS FINALES

Iniciamos este capítulo retomando algunas definiciones y resultados de la TCT. Esperamos que la introducción a la Teoría G en este capítulo haya dejado claro al lector que la Teoría G no contradice la TCT, sino que la extiende y complementa. Comparando la ecuación básica ($X = T + E$) de la TCT con la descomposición de la puntuación observada en las Ecuaciones 3 y 9, directamente muestra que el puntaje universo ($\mu_p \equiv \mu + \alpha_p$) en la Teoría G conceptualmente corresponde con la puntuación verdadera (T) en la TCT y que los demás componentes en las Ecuaciones 3 y 9 se juntan en la puntuación error (E). Asimismo, hay correspondencias claras entre la descomposición de la varianza de las puntuaciones observadas en la TCT (Ecuación 2) y la Teoría G (Ecuaciones 4 y 10), así como entre la definición de la confiabilidad en la TCT (Ecuación 3) y los coeficientes de generalizabilidad (en las Ecuaciones 7, 8, 11 y 12). Lo anterior muestra que la ventaja principal de la Teoría G es que separa la influencia de las distintas fuentes de error en la medición. En general, los errores son inevitables y muy variados, pero conviene mirarlos como oportunidades que nos permitan aprender y mejorar nuestros esfuerzos, en el mismo espíritu que Thomas Edison expresó en su frase célebre mostrada al inicio del capítulo. La Teoría G ofrece las herramientas para conocer mejor los errores y, aprovechando la información proporcionada por los análisis, obtendremos mediciones más precisas y tomaremos decisiones más acertadas.

Algunos autores también han comentado sobre las relaciones entre la Teoría G y el enfoque psicométrico de la teoría de respuesta al ítem (TRI, véase el capítulo anterior). El esfuerzo más notable viene de Briggs y Wilson (2007), quienes combinan los dos enfoques en un nuevo modelo: GIRM (*Generalizability in Item Response Modeling*). Al lector interesado en conocer este y otros esfuerzos de conciliar la Teoría G y la TRI, le puede resultar útil la tesis doctoral de Choi (2013).

Agradecimiento

Los autores agradecen a José J. Naveja por la lectura crítica y comentarios realizados sobre una versión previa de este capítulo.

REFERENCIAS

Brennan, R. L. (2001). *Generalizability theory.* Springer.

Briggs, D. C., y Wilson, M. (2007). Generalizability in item response modeling. *Journal of Educational Measurement*, *44*(2), 131–155. doi: https://doi.org/10.1111/j.1745-3984.2007.00031.x

Choi, J. (2013). *Advances in combining generalizability theory and item response theory* [Tesis de doctorado no publicada]. Universidad de California en Berkeley. doi: 10.13140/RG.2.1.4458.1285

Cronbach, L. J., Gleser, G. C., Nanda, H., y Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability of scores and profiles.* Wiley.

Cronbach, L. J., Rajaratnam, N., y Gleser, G. C. (1963). Theory of generalizability: A liberation of reliability theory. *The British Journal of Statistical Psychology*, *16*(2), 137–163. doi: 10.1111/j.2044-8317.1963.tb00206.x

Dizon, S., Malcolm, J. C., Rethans, J.J., y Pugh, D. (2021). Assessing the validity of an OSCE developed to assess rare, emergent or complex clinical conditions in endocrinology & metabolism. *BMC Medical Education*, *21*, 288. doi: 10.1186/s12909-021-02653-4

Espinosa-Vázquez, O., Martínez-González, A., Sánchez-Mendiola, M., y Leenen, I. (2017). Análisis de un examen clínico objetivo estructurado en odontología desde la teoría de la generalizabilidad. *Investigación en Educación Médica*, *6*(22), 109–118. doi: 10.1016/j.riem.2016.09.001

Pardo, A., y San Martín, R. (2010). *Análisis de datos en ciencias sociales y de la salud II*. Síntesis.

Tejedor, F. J. (2019). *Análisis de varianza: Introducción conceptual y diseños básicos* (Cuadernos de Estadística no. 3, 3ra ed.). La Muralla.

Trejo-Mejía, J. A., Sánchez-Mendiola, M., Méndez-Ramírez, I., y Martínez-González, A. (2016). Reliability analysis of the objective structured clinical examination using generalizability theory. *Medical Education Online*, *21*(1), 31650. doi: 10.3402/meo.v21.31650

Webb, N. M., Shavelson, R. J., y Haertel, E. H. (2006). Reliability coefficients and generalizability theory. *Handbook of Statistics*, *26*, 81–124. doi: 10.1016/S0169-7161(06)26004-8

Capítulo 19

INTRODUCCIÓN A LA GENERACIÓN AUTOMÁTICA DE ÍTEMS

Eduardo Backhoff Escudero

"El propósito de la educación es formar individuos que se puedan realizar como seres humanos y que se conviertan en ciudadanos capaces de contribuir a construir un país justo y próspero; el objetivo de la evaluación es verificar y coadyuvar a que este propósito se cumpla."

INTRODUCCIÓN

Todo proceso educativo busca alcanzar la meta final de formar a los individuos para que desarrollen sus capacidades, se realicen como seres humanos y se conviertan en ciudadanos productivos que contribuyan a mejorar el país donde viven. Independientemente del modelo pedagógico que se utilice, la evaluación es un componente consustancial del proceso educativo, pues representa la forma idónea de comprobar en qué medida las metas de aprendizaje se han alcanzado por cada uno de los estudiantes y, con base en esta información, retroalimentar su ejecución durante el curso escolar y certificar la adquisición de sus habilidades y conocimientos al final de este.

Aunque los exámenes escolares no son la única manera de evaluar las competencias de un estudiante, sí representan la forma más común y práctica de hacerlo. En el caso de la educación básica, los docentes se preparan en las escuelas normales para enseñar y evaluar a sus estudiantes (aunque con muchas limitaciones). En los niveles de educación media superior (EMS) y superior (ES) los profesores se forman en la práctica de la docencia; ya que se trata de profesionistas que no se formaron para ejercer la pedagogía. De esta manera, por lo general, sus actividades de enseñanza y de evaluación no se fundamentan en ninguna teoría del aprendizaje; se trata de prácticas didácticas que, por intuición o imitación, se ejercitan y se mejoran a través del método de ensayo y error. En algunos casos se obtienen buenos resultados y en muchos otros solo se logran prácticas pedagógicas mediocres.

Independientemente de su preparación o experiencia pedagógica, todos los docentes tienen la necesidad de preparar clases, elaborar materiales didácticos y utilizar exámenes para evaluar el logro académico de sus estudiantes. En algunas instituciones educativas se utilizan exámenes departamentales o, bien, exámenes de egreso para certificar las competencias adquiridas, como es el caso del examen de las competencias médicas (UNAM, 2021). En

estos casos, donde se evalúa a una cantidad importante de estudiantes de manera repetida, es común que las instituciones elaboren bancos de preguntas (ítems o reactivos) para construir diferentes versiones de una prueba y así evitar que los estudiantes o egresados conozcan las preguntas de exámenes previos; en cuyo caso, los resultados de los exámenes perderían su validez.

La elaboración de exámenes de gran escala[1] requiere del trabajo colegiado de expertos formados por docentes del nivel educativo de que se trate, especialistas en enseñanza de la disciplina correspondiente, expertos en evaluación del aprendizaje y estadísticos especializados en instrumentos de evaluación (Tiana, 1996). La limitación que presenta este trabajo es que los reactivos elaborados se desgastan una vez que se han utilizado en forma masiva o en repetidas ocasiones, por lo que se necesita construir nuevas versiones de preguntas y, en el mejor de los casos, combinarlas con versiones anteriores para hacer rendir el alto costo de la elaboración de este tipo de exámenes. Ante la necesidad de contar con grandes bancos de reactivos, desde los años sesenta, se han hecho diferentes intentos para poder automatizar (y, así, abaratar los costos y facilitar los procesos) la manera de construir una gran cantidad de preguntas; campo al que hoy se le conoce como Generación Automática de Reactivos (GAI) (Irvine y Kyllonen, 2002).

Por su relevancia para el buen funcionamiento de las instituciones educativas, el propósito de este capítulo es introducir al profesor universitario y a las autoridades encargadas de las evaluaciones institucionales en el tema de la GAI, a fin de que evalúen su posible utilización en el quehacer de su competencia. Para lograr este objetivo, se narrará brevemente los antecedentes históricos del nacimiento y evolución de la GAI, se describirán y ejemplificarán los principales modelos de estos generadores de reactivos, se explicará la forma en que el advenimiento de las computadoras en el sector educativo ha hecho posible avanzar sustancialmente en el campo de la GAI y se ejemplificará el uso de esta tecnología con el Generador Automático de Exámenes (GenerEx) (Ferreyra y Backhoff, 2016; Sánchez y Backhoff, 2015); sistema que fue desarrollado en México por Métrica Educativa A.C. (www.metrica.edu.mx) y que se utiliza para generar los reactivos de varios exámenes nacionales, entre los que se encuentra el Examen de Competencias Básicas (Excoba) (Backhoff, Larrazolo, Ramírez y col, 2015).

ANTECEDENTES DE LA GAI

Los orígenes para evaluar las diferencias individuales de las personas se pueden ubicar en los inicios del siglo pasado, con los trabajos de Alfred Binet sobre la medición de la inteligencia que se utilizó, en un principio en Francia, para identificar a los estudiantes que podrían presentar dificultades para aprender en la escuela (Beltrán-Llera y Pérez-Sánchez, 2011). Muy pronto, en la Primera Guerra Mundial, los tests de inteligencia se empezaron a utilizar en el

[1] También conocidas como estandarizadas, son aquellas que se utilizan en cientos o miles de estudiantes, como es el caso de los exámenes de admisión o certificación.

sector militar para identificar aquellos reclutas que tuvieran las capacidades necesarias para incorporarse a los trabajos de la milicia (Ben-Simon y Cohen, 2004). Igualmente, estas pruebas se empezaron a utilizar masivamente en el sector educativo de los Estados Unidos, con el objeto de poder ubicar a los estudiantes en los distintos cursos escolares de acuerdo con sus capacidades intelectuales (Ben-Simon y Cohen, 2004).

Las evaluaciones de gran escala para medir el logro escolar de los estudiantes se popularizaron a partir de los años cincuenta del siglo pasado, lo que impulsó el nacimiento y crecimiento exponencial de empresas dedicadas a la elaboración de todo tipo de pruebas de aprendizaje, ya sea para seleccionar a los aspirantes de una institución, diagnosticar sus habilidades y conocimientos escolares o, bien, certificar las competencias profesionales (Ben-Simon y Cohen, 2004).

Algunas instituciones educativas empezaron a utilizar pruebas departamentales estandarizadas, al darse cuenta que los estudiantes acreditaban los cursos de distintas asignaturas con un dominio de estas muy diferente. En otras instituciones se implementaron pruebas de egreso de ciertos niveles educativos para poder obtener el certificado de terminación de estudios correspondiente; tal es el caso del examen terminal del bachillerato en Francia, *Baccalauréat* (Wikipedia, 2021). En todos los casos, fue notorio que para que las evaluaciones cumplieran su función estas deberían ser desarrolladas con altos estándares de calidad, para lo cual era necesario contratar a especialistas en el desarrollo de pruebas de aprendizaje. Sin embargo, una vez que se utilizaban las pruebas se desgastaban rápidamente, dado que no había manera de evitar que sus contenidos se filtraran y se dieran a conocer entre los estudiantes que aún no se evaluaban.

La solución a este problema fue, inicialmente, la elaboración de bancos de reactivos para formar distintas versiones de un examen. Sin embargo, si bien se solucionaba el problema de contar con preguntas distintas de una versión a otra, se creaba un problema nuevo: el tener que utilizar exámenes con distintos niveles de dificultad, pues es imposible garantizar *a priori* que dos preguntas que evalúan un mismo contenido escolar tengan el mismo nivel de dificultad; condición que se extiende al examen en su totalidad. Por ello, era necesario utilizar distintos procedimientos para igualar el nivel de dificultad de dos versiones de una misma prueba, lo que se conoce en la literatura especializada como equiparación (*equating*, en inglés). Lo anterior, también abonó a que se incrementaran los costos de la elaboración de exámenes. Teniendo la necesidad de elaborar una cantidad de exámenes y versiones distintas de cada uno de ellos, surgió la idea de poder automatizar la elaboración de sus reactivos; lo que resolvería el problema, al menos, en forma parcial. Posiblemente, el primer intento publicado en este sentido fueron los trabajos de Osburn (1968) y Hively, Patterson y Page (1968), quienes propusieron el uso de *formas de ítems*. Este mecanismo para generar decenas de reactivos consistía en la redacción de la estructura sintáctica de una pregunta, que permitía poder reemplazar algunos de sus elementos, con la finalidad de contar con varias versiones para evaluar un mismo contenido curricular. El Recuadro 1 presenta un ejemplo típico de esta técnica.

Recuadro 1. Ejemplo de una forma de ítem, de acuerdo con Osburn (1968) y Hively, Patterson y Page (1968).

Texto fijo de la pregunta
Una costurera borda ______ manteles en _____ horas. Si tiene que bordar un juego de______ manteles, ¿cuántas horas tiene que trabajar para terminarlo?

Elementos intercambiables (A1, A2 y A3)
Una costurera borda __ **A1**___ manteles en __ **A2**__ horas. Si tiene que bordar un juego de___ **A3**___ manteles, ¿cuántas horas tiene que trabajar para terminarlo?

Reglas de los elementos intercambiables
A1: la variable puede tomar valores en un rango de 5 a 50
A2: la variable puede tomar valores en un rango de 25 a 250
A3: la variable puede tomar valores en un rango de 3 a 30

Fuente: Adaptación de Sánchez y Backhoff (2015).

Como se aprecia en el recuadro, para generar versiones equivalentes de reactivos se establecían reglas que limitaban la dificultad de las distintas versiones de las preguntas. Estas reglas se referían a los valores mínimos y máximos de cada uno de los elementos intercambiables del reactivo, o bien del tipo de operación matemática que debería de realizar el estudiante para resolver los problemas planteados (Hively, Patterson y Page, 1968).

Veinte años después, Haladyna y Shindoll (1989) propusieron la elaboración de "moldes de reactivos" (*Item Shell*, en inglés) con la idea de facilitarles el trabajo a los redactores de ítems y poder realizar su tarea con mayor eficiencia y rapidez. Estos moldes contienen la estructura semántica y otros elementos de los ítems convencionales. Es decir, se trata de ítems "huecos" cuya base del reactivo ya estaba redactada, con excepción de algunos elementos relacionados con su contenido, como se muestra en el recuadro 2.

Recuadro 2. Ejemplo de un molde de ítem siguiendo la propuesta de Haladyna y Shindoll (1989)

Base del reactivo original

¿Cuál es el objeto de estudio de la Ecología?

- Las relaciones de los organismos y su medio.
- La estructura y función de los seres vivos.
- La contaminación atmosférica.
- Los deterioros del medio ambiente.

Molde de reactivo

¿Cuál es el objeto de estudio de ______________________?

Fuente: Adaptación de Sánchez y Backhoff (2015).

Los moldes de ítems, por lo general, se seleccionan de bancos de reactivos que han sido probados y validados con anterioridad y de los cuales se conoce su funcionamiento métrico (o estadístico), como son sus niveles de dificultad. La idea central de estos moldes es ahorrarles tiempo a los elaboradores de reactivos, quienes deben de concentrarse en las competencias a evaluar y no en la estructura semántica, gramatical o sintáctica que requiere una pregunta. De esta manera se pueden crear múltiples versiones de un ítem con el que se desea evaluar un mismo conocimiento o habilidad.

Solano-Flores, Shavelson y Schneider (2001) avanzaron en el diseño de moldes de ítems acuñando el término "plantilla" (en inglés, *template*), para referirse a un conjunto de instrucciones que sirven para desarrollar ejercicios evaluativos de bajo costo y en poco tiempo. Esto permite cuidar los aspectos de redacción, gramática y ortografía a partir de una estructura sintáctica preestablecida, cuya función principal es facilitar que el elaborador del ítem atienda los contenidos a evaluar y no se distraiga en su redacción. Sin embargo, como los autores lo anticipan, este procedimiento no garantiza que los reactivos que se elaboren sean métricamente equivalentes. El recuadro 3 muestra un ejemplo de un molde de ítem con una estructura sintáctica base (fija) y una serie de especificaciones para rellenar los espacios en el texto (que se señalan entre paréntesis y en itálicas). En la parte inferior del recuadro se muestra un ítem generado con base en este modelo, donde en subrayado se identifican los elementos que conforman el nuevo ítem.

Recuadro 3. Ejemplo de un molde de ítem y del reactivo resultante

Molde de reactivo con estructura sintáctica

(Tema o contenido a evaluar)

La evidencia científica indica que (*presentar un personaje, elemento o nombre de un tema central o fenómeno principal*) ha contribuido en gran medida a (descripción breve del tema central). Utilizando tus conocimientos acerca de (tema central) y del concepto (contenido asociado al tema central):

- Describe cómo (*elementos 1, 2 y 3 asociados al tema central como evidencia del mismo*) se relacionan con el fenómeno de (tema central).
- Explica por qué decir: *(afirmación o enunciado que indique una idea errónea acerca del tema central*), es una afirmación errónea.
- Explica por qué (*situación concreta asociada al tema central*) es provocado por la relación entre (*elementos 4, 5, 6 y 7 relacionados como causas del tema central*).

Tus respuestas deben mostrar un dominio preciso y profundo del conocimiento de los conceptos, principios y razonamientos relacionados con el (tema central).

Ítem resultante del molde del reactivo

El calentamiento global

La evidencia científica indica que la especie humana ha contribuido en gran medida a la presencia del calentamiento global. Utilizando los conocimientos del tema calentamiento global y del concepto huella ecológica:

- Describe cómo el aumento en los niveles del mar, el derretimiento de las capas polares y la desaparición de muchas especies de animales y plantas se relacionan con el fenómeno del calentamiento global.
- Explica por qué decir: "el calentamiento global es un proceso natural del planeta", es una afirmación errónea.
- Explica por qué el aumento en la temperatura promedio del planeta es provocado por la relación entre el uso indiscriminado de combustibles fósiles, el uso desmesurado de fertilizantes, los procesos industriales y la pérdida de bosques.

Tus respuestas deben mostrar un dominio preciso y profundo del conocimiento de los conceptos, principios y razonamientos relacionados con el calentamiento global.

Nota: En paréntesis y en itálicas se señalan los elementos del molde del reactivo que se deben cambiar y en subrayado los elementos que conforman un reactivo, de acuerdo con dicho molde.

Fuente: Adaptado de Sánchez y Backhoff (2015).

El advenimiento de las computadoras y su impacto en la GAI

En las últimas décadas, pocos temas han sido tan relevantes en el mundo como el arribo de la informática, disciplina que ha impactado en todos los ámbitos de la humanidad, entre ellos la educación. Muestra de este impacto es el uso de las tecnologías de la información para atender los efectos del COVID-19, que obligó a las escuelas a cerrar sus puertas y adoptar la modalidad de educación a distancia; lo que requirió que los estudiantes aprendieran a aprender en casa.

Específicamente, en el campo de la educación, los recursos digitales han posibilitado la innovación de nuevas formas de evaluar el aprendizaje, tales como: los exámenes asistidos por computadora, las pruebas adaptativas, los simuladores para evaluar competencias profesionales y el desarrollo de la GAI. Las ventajas que ofrece la evaluación por medios digitales, en comparación con las de lápiz y papel, son numerosas. Entre ellas podemos destacar las siguientes:

- Se generan preguntas y versiones de exámenes de manera automática.
- Se mejora la presentación de los exámenes al utilizar diversos medios digitales, tales como ilustraciones, fotografías, audios, animaciones y videos.
- Se califican las respuestas de los estudiantes de forma automática e inmediata.
- Se generan reportes individualizados de resultados de manera eficiente.
- Se mejora la seguridad de los contenidos de los exámenes.
- Se transparentan los procesos de evaluación y se mejora la imagen institucional.

Una ventaja especial de la evaluación computarizada es que permite utilizar preguntas distintas a las de opción múltiple, cuyas respuestas son más naturales o "auténticas", como sería la escritura de ecuaciones algebraicas, la ubicación de puntos cartesianos en un plano, el balance de fórmulas químicas, el llenado de tablas numéricas, el subrayado de oraciones en un párrafo, la categorización de conceptos, etcétera.

Por estas y otras ventajas, prácticamente, todas las pruebas más prestigiadas del mundo se administran a través de plataformas digitales, ya sea local o remotamente, como son los casos de las pruebas internacionales de PISA (Programme for International Student Assessment), TOEFL (Test of English as a Foreign Language), SAT (Scholastic Assessment Test) y GRE (Graduate Record Examinations).

Por otra parte, el uso de la tecnología digital también ha ayudado a que se avance sustancialmente en el desarrollo de la GAI que, hasta fines del siglo pasado, era muy limitado. La publicación del libro *Item Generation for Test Development* (Irvine y Kyllonen, 2002), que presenta una compilación de las aportaciones más importantes de la GAI a lo largo de su historia, representa un parteaguas en este campo de la evaluación. En este libro se documenta que solo con el apoyo de las computadoras personales se puede aspirar a desarrollar generadores de reactivos de una manera más creativa, sofisticada y eficiente, que responda a las necesidades de quienes se dedican a elaborar instrumentos de evaluación de gran escala.

Las nuevas aproximaciones de la GAI requieren que se desarrollen modelos de ítems que permitan elaborar y manipular los distintos componentes de un reactivo, de tal manera que sea posible medir la competencia de las personas de manera equivalente e intercambiable (Bejar, 2002; Bejar, Lawless, Morley et al., 2003; LaDuca, Staples, Templeton et al., 1986). Un problema de la GAI es que la modificación de cualquier elemento o componente de un ítem, potencialmente, puede cambiar la esencia de la competencia que se pretende medir (conocimiento, habilidad o actitud) o, bien, la dificultad con la que se le mide. Por ello, es importante saber que hay dos aproximaciones teóricas para elaborar instrumentos de evaluación mediante la GAI: la teoría fuerte y la teoría débil. Quienes utilizan la teoría fuerte parten de los principios de teorías psicológicas capaces de explicar los procesos cognitivos que utilizan las personas para poder responder correctamente a una pregunta (Gitomer y Bennett, 2002), con lo que se pueden manipular las propiedades y dificultad de los reactivos generados (Gierl y Lai, 2012) gracias al sustento teórico que los soporta (Lai, Alves y Gierl, 2009). El problema de esta aproximación es que no existen, por ahora, suficientes teorías cognoscitivas para llevar a cabo estos principios a la práctica escolar en los distintos ámbitos del conocimiento (Gitomer y Bennett, 2002).

Por su parte, el uso de la teoría débil de la GAI no requiere que se precisen los procesos cognoscitivos necesarios para responder un reactivo. Pero sí es importante asegurar que las versiones de los ítems que se generen sean invariantes –es decir, que no cambien sus características fundamentales de una población a otra– para poder producir reactivos equivalentes, tanto conceptualmente como métricamente (Drasgow, Luecht y Bennett, 2006). A esta propiedad se le conoce en la literatura especializada con el nombre de isomorfismo, propiedad que se tiene que probar empíricamente (Gierl y Lai, 2012).

Por lo general, los instrumentos que se utilizan en el ámbito educativo, en los que se evalúan dominios amplios del conocimiento –como en los exámenes de admisión, departamentales o de certificación– no requieren que se conozcan los procesos cognoscitivos exactos (como en el caso de la teoría fuerte), pero sí es necesario generar grandes cantidades de reactivos isomorfos. Esto requiere que los modelos de ítems contengan los elementos bien definidos que se incluirán en una tarea evaluativa y que serán intercambiados para elaborar distintas versiones de los ítems, a saber: 1) la base del reactivo, 2) los elementos sustituibles y 3) las reglas para generar ítems equivalentes (como se describirá y ejemplificará en el siguiente apartado).

GENERADOR AUTOMÁTICO DE EXÁMENES (GENEREX)

El GenerEx[2] se puede considerar como un prototipo de la GAI, ya que su principal función es generar de manera automática una gran cantidad de reactivos isomorfos, para poder construir una diversidad de versiones de exámenes equivalentes. Dado que este sistema informático se fundamenta en la teoría débil de los generadores, hay que comprobar que las distintas versiones de reactivos que genera, no solo sean conceptualmente equivalentes –lo que se puede hacer a través de la opinión de expertos– sino que, además, sus características métricas (al menos, su dificultad) sean semejantes, lo que solo se puede comprobar a través de estudios empíricos (Ferreyra y Backhoff, 2016).

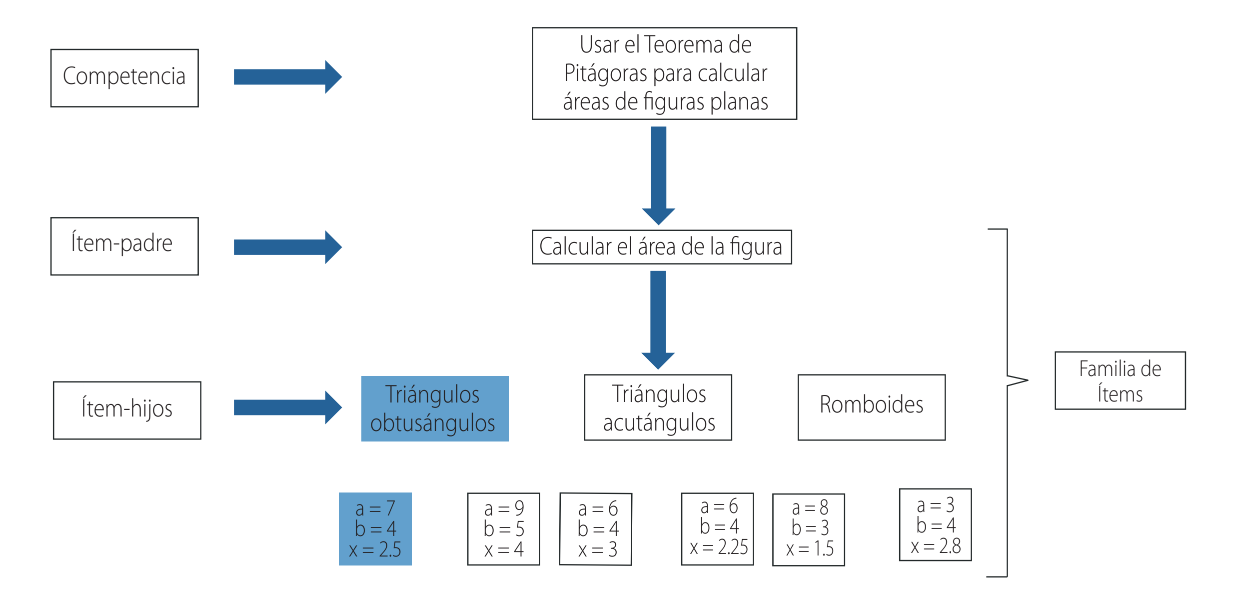
El GenerEx se terminó de desarrollar en 2013, cinco años después de que inició su diseño (Backhoff, Larrazolo y Tirado, 2013). Su antecedente inmediato es el Sistema Computarizado de Exámenes (SICODEX) que se desarrolló, en 1993, con el propósito de administrar exámenes de opción múltiple por computadora (Backhoff, Ibarra y Rosas, 1996), como fueron los casos del Examen de Habilidades y Conocimientos Básicos (EXHCOBA) y del Examen de Egreso del Idioma Inglés del nivel intermedio (EXEDII) (Velazco, Anguiano y Larrazolo, 2007).

Para superar las limitaciones del formato de opción múltiple, avanzar en el uso de la evaluación asistida por computadora y facilitar la generación automática de reactivos, en 2008 se inició el desarrollo de manera conjunta del GenerEx y del Examen de Competencias Básicas (Excoba). Con estos nuevos desarrollos se buscó alcanzar tres metas: evaluar las competencias básicas que se definen en los planes y programas de estudio de la educación obligatoria mexicana; evaluar estas competencias de la manera lo más auténtica posible, es decir, procurando que los estudiantes, en vez de seleccionar las respuestas, las construyeran; y, generar una infinidad de reactivos isomorfos y, en consecuencia, de versiones de exámenes equivalentes.

Para entender cómo funciona el GenerEx, hay que conocer en qué consiste una familia de ítems. Para ello, se muestra el siguiente ejemplo. Supóngase que se desea evaluar el dominio de la competencia de utilizar el Teorema de Pitágoras para calcular áreas de figuras planas (por ejemplo, triángulos). Esta competencia (del nivel de educación básica) puede evaluarse de muchas maneras, pero se requiere especificar las reglas para construir los reactivos con los que se evaluará. Es decir, hay que especificar la familia de reactivos correspondiente, que consta de tres componentes: la competencia a evaluar, así como las características del ítem-padre y de los ítems-hijo. La Figura 1 muestra un ejemplo de estos tres componentes.

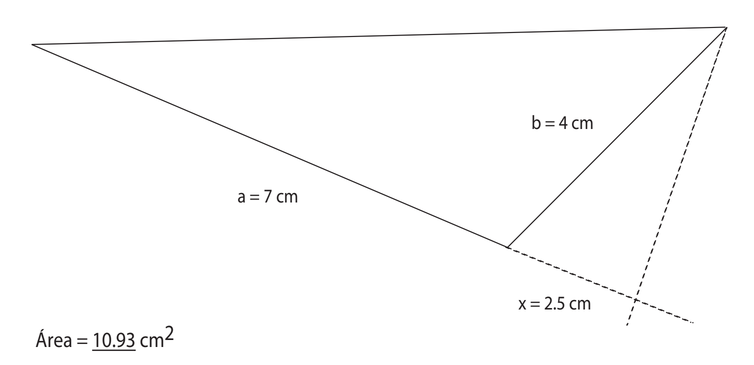
[2] Con registro en el Instituto Nacional de Derechos de Autor.

Figura 1. Familia de reactivos de la competencia uso del Teorema de Pitágoras para calcular perímetros y áreas de figuras planas



En la Figura 2 se muestra el ejemplo de un reactivo generado con la información del modelo de ítem descrito (que en la Figura 1 se resalta sombreándolo). Utilizando la información proporcionada en la Figura 2 (variables a, b y x), el estudiante debe calcular el área del triángulo, para lo cual tiene que calcular, primero, la longitud de la línea punteada con el uso del Teorema de Pitágoras.

Figura 2. Ejemplo de un ítem-hijo para evaluar la competencia de calcular el área de una figura plana regular, utilizando el Teorema de Pitágoras



Es importante mencionar que una familia de reactivos forma parte de un modelo de ítem, que contiene los siguientes componentes: la definición de la competencia a evaluar, la estrategia para evaluarla (en este caso, presentar triángulos o romboides con información incompleta), las reglas para combinar los elementos de la familia de reactivos (en este caso, los valores que pueden tomar las variables a, b y x), el tipo de ejecución que se le solicitará al estudiante (en este caso, escribir la cifra del cálculo correspondiente) y la forma en que se calificarán las respuestas (como podría ser un margen de error aceptable, que en este caso no está especificado). Finalmente, un modelo de ítem puede solicitar más de una respuesta al estudiante; en el ejemplo anterior, se podría haber solicitado también el área del triángulo formado por las líneas punteadas.

A la fecha, el GenerEx se ha utilizado, junto con el Excoba, en procesos de admisión de varias instituciones de educación media superior y superior mexicanas y del extranjero. Igualmente, se ha probado con diversos exámenes diagnósticos de inglés y de matemáticas. La experiencia ha mostrado que el GenerEx es un sistema muy eficiente para generar reactivos y exámenes equivalentes, que se administran y se califican por medios computacionales. Como ya se señaló anteriormente, es necesario verificar que las distintas versiones de reactivos sean isomorfos o métricamente equivalentes, como lo muestran algunos estudios realizados con el Excoba (ver, por ejemplo, Ferreyra y Backhoff, 2016).

REFLEXIONES FINALES

El propósito de este capítulo fue explicar a los docentes, así como a los responsables de los departamentos de evaluación de las instituciones educativas, en qué consisten los generadores automáticos de reactivos, de dónde y desde cuándo proviene el interés por utilizarlos, cómo han ido evolucionando a lo largo del tiempo y qué impacto ha tenido el advenimiento de la tecnología digital en el campo de la evaluación educativa y en el área de la GAI, en particular. Para lograr el propósito del capítulo, se describe el funcionamiento de un generador automático de ítems desarrollado en México: el GenerEx.

Dos aspectos que se deben resaltar de la GAI son sus ventajas y limitaciones. Entre las primeras destacan la eficiencia con que se pueden generar reactivos y exámenes equivalentes, que evita el desgaste que tienen las pruebas de gran escala, debido a su uso intensivo y a la facilidad con que se pueden socializar sus contenidos y, por ello, invalidar sus resultados. Una aportación específica del GenerEx al campo de la evaluación radica en la posibilidad de utilizar reactivos distintos a los de opción múltiple y, con ello, hacer más auténtica la evaluación del aprendizaje; como sería la escritura de una ecuación matemática, el balanceo de una ecuación química o la ubicación de coordenadas geográficas en un mapa.

Por otro lado, dos de las grandes limitaciones que presenta la GAI tienen que ver con su dificultad para desarrollar y probar el isomorfismo de los reactivos que generan. Respecto a la primera limitación, es importante subrayar que la elaboración de modelos de ítems que requiere la GAI es una tarea compleja, en la que se debe tener cuidado en el nivel de detalle con que se precisan las instrucciones para generar ítems. Es necesario contar con

procedimientos rigurosos que garanticen que la forma en que se generan los reactivos a partir de un mismo modelo son lo suficientemente rigurosos teórica y técnicamente como para garantizar, al menos, su equivalencia conceptual. En este sentido una segunda limitación de los generadores como el GenerEx, que se basan en la teoría débil, se refiere a que la semejanza conceptual entre dos reactivos no garantiza su equivalencia métrica, por lo que es necesario que cada ítem se pilotee para conocer su nivel de dificultad (y otras propiedades métricas) y poder sustentar su isomorfismo. Una solución a este problema es equiparar[3] las versiones de un examen, partiendo de la premisa de que las versiones son semejantes conceptualmente, pero no idénticas psicométricamente.

Con todas sus fortalezas y limitaciones, el desarrollo del GenerEx y su uso, a partir de 2013, en diversas instituciones públicas y privadas de educación superior y media superior, marca un parteaguas de la evaluación asistida por computadora en México y en una de sus innovaciones más recientes: la GAI.

En resumen, la GAI es un campo emergente de la evaluación asistida por computadora que promete revolucionar la evaluación del aprendizaje a gran escala. Esto es especialmente cierto para las instituciones educativas que requieren elaborar o utilizar frecuentemente distintas evaluaciones con propósitos de ingreso, diagnóstico y certificación o, bien, que desean implementar exámenes departamentales en distintas áreas del conocimiento. Sería deseable que en México se impulsara el desarrollo e innovación en este campo de la evaluación educativa en aquellas instituciones educativas que utilicen exámenes estandarizados o de gran escala.

REFERENCIAS

Backhoff, E., Ibarra, M. A. y Rosas, M. (1996). Desarrollo y validación del sistema computarizado de exámenes (SICODEX). *Revista de la Educación Superior*, Vol. XXXV, No. *1*(97), pp. 41-54.

Backhoff, E., Larrazolo, N., Ramírez, J.L., Rosas, M., y Tirado, F. (2015). Excoba: Examen de Competencias Básicas. México: Instituto Nacional de Derechos de Autor.

Backhoff, E., Larrazolo, N., Tirado, F., (2013 noviembre). Desarrollo y validación de un Generador Automático de Reactivos de respuesta construida para elaborar exámenes computarizados de ingreso a la educación superior. Memoria III Clabes Conferencia Latinoamericana sobre el Abandono en la Educación Superior.

Beltrán-Llera, J. y Pérez-Sánchez, L. (2011). Más de un siglo de psicología educativa. Valoración general y perspectivas de futuro. *Papeles del Psicólogo*, *32*(3), 204-231.

Bejar, I. (2002). Generative testing: from conception to implementation. En S. H. Irvine y P. C. Kyllonen (eds.), Item generation for test development (pp. 199-218). Mahwah, NJ: Lawrence Erlbaum Associates.

[3] La equiparación de dos versiones de un examen consiste en ajustar sus puntuaciones para que tengan el mismo nivel de dificultad.

Bejar, I., Lawless, R., Morley, M., Wagner, M., Bennett, R., y Revuelta, J. (2003). A feasibility study of on-the-fly item generation in adaptive testing. *Journal of Technology, Learning, and Assessment, 2*(3), 1-29.

Ben-Simon, A., y Cohen, Y. (2004). International assessments: Merits and pitfalls. Trabajo presentado en la 30ª Conferencia Anual de la Asociación Internacional de Medición Educativa. Filadelfia, PA.

Drasgow, F., Luecht, R., y Bennett, R. (2006). Technology and testing. En R. L. Brennan (e.), Educational measurement (pp. 471-516). Washington, DC: American Council on Education.

Ferreyra, F., y Backhoff, E. (2016). Validez del Generador Automático de Ítems del Examen de Competencias Básicas (Excoba). RELIEVE, *Revista Electrónica de Investigación y Evaluación Educativa, 22*(1) 1-16. Recuperado de http://www.uv.es/RELIEVE/v22n1/RELIEVEv22n1_2.htm

Gierl, M., y Lai, H. (2012). Using weak and strong theory to create item models for automatic item generation: Some practical guidelines with examples. En M. J. Gierl & T. Haladyna (eds.). Automatic item generation: Theory and practice. New York: Routledge.

Gitomer, D., y Bennett, R. (2002). Unmasking Constructs Through New Technology, Measurement Theory, and Cognitive Science (Memorandum de investigación, febrero de 2002, RM-02-01). Educational Testing Service. Statistics and research division. Princeton, NJ.

Haladyna, T. M. y Shindoll, R. R. (1989). Item shells: A method for writing effective multiple-choice test items. Evaluation and the Health Professions, 12, 97-106.

Hively, W., Patterson, H., y Page, S. (1968). A "universe-defined" system of arithmetic achievement tests. *Journal of Educational Measurement,* (5), 275-290.

Irvine, S., y Kyllonen P. (eds.). (2002). Item generation for test development. Mahwah, NJ: Lawrence Erlbaum Associates.

LaDuca, A., Staples, W., Templeton, B., y Holzman, G. (1986). Item modeling procedure for constructing content-equivalent multiple-choice questions. *Medical Education,* (20), 53-56.

Osburn, H. G. (1968). Item sampling for achievement testing. Educational and psychological measurement, 28, 95-104.

Sánchez, C., y Backhoff, E. (2015). Generación automática de ítems: una nueva aproximación para evaluar el aprendizaje (Una revisión). REVALUE, *Revista de Evaluación Educativa,* 4(2), 1-25

Solano-Flores, G., Jovanovic, J. Shavelson, R. J. y Bachman, M. (1999). On the development and evaluation of a shell for generating science performance assessment. *International Journal of Science Education, 21*(3), pp. 293-315.

Solano-Flores, G., Shavelson, R. J. y Schneider, S. A. (2001). Expanding the notion of assessment shell: from task development tool to instrument for guiding the process of science assessment development. *Revista Electrónica de Investigación Educativa, 3*(1). Recuperado de http://redie.uabc.mx/vol3no1/contents-solano.html

Tiana, A. (1996). La evaluación de los sistemas educativos. *Revista Iberoamericana de Educación,* 10, 37-61.

Universidad Nacional Autónoma de México [UNAM] (2021). Examen de Competencia Académica Ciclo 2021-2022. Recuperado de: https://www.fmposgrado.unam.mx/index.php/examen-de-competencia-academica-ciclo-2021-2022

Velazco, V., Anguiano, M.L. y Larrazolo, N. (2007). Propuesta metodológica para la formulación de una conceptualización del constructo de un examen de certificación del inglés como lengua extranjera. *Revista Electrónica de Investigación Educativa, 6*(2), 1-22. Recuperado de: http://redie.uabc.mx/vol9no2/contenido-velasco.html

Wikipedia (2021). *Baccalauréat* Recuperado de: https://en.wikipedia.org/wiki/Baccalaur%C3%A9at

Capítulo 20

EQUIPARACIÓN EN LAS PRUEBAS DE DESEMPEÑO: ALGUNAS TÉCNICAS MÁS COMUNES EN LA EVALUACIÓN EDUCATIVA

Rafael Enrico Sánchez Mayorga, Laura Delgado Maldonado

"La calidad no es un acto, sino un hábito."
Aristóteles

INTRODUCCIÓN

Las universidades y las distintas instancias de evaluación educativa cuentan con programas de evaluación que implican una variedad de recursos sea para mejorar el aprendizaje de los estudiantes o para medir su desempeño académico, haciendo inferencias válidas a partir de los resultados de las evaluaciones con las que cuentan; por ejemplo, para revisar el plan de estudios; como un diagnóstico para informar a los estudiantes de su progreso; informar al público sobre el desempeño escolar; para ser utilizado como guía en la toma de decisiones sobre estudiantes, maestros o escuelas; para el ingreso, egreso o certificación de un nivel académico proporcionando información útil que apoye las inferencias de los resultados.

Por lo que, cada vez es necesario establecer y mantener la calidad técnica de un programa de pruebas, ya que esta impacta directamente en la validez de las evaluaciones y el grado en que la evidencia y la teoría apoyan las interpretaciones de los puntajes de las pruebas (AERA, APA y NCME, 2014). Una de las técnicas que apoya a alcanzar este objetivo es la equiparación (*Equating),* la cual es un método que se lleva a cabo para establecer puntuaciones comparables entre las diferentes formas de una prueba, lo cual permite que se usen indistintamente en un periodo de tiempo determinado.

En el diseño de una prueba nos preguntamos, porqué es importante contar con más de una forma para su aplicación o porqué es necesario llevar a cabo la equiparación entre las distintas formas que se aplican de un mismo instrumento a través del tiempo; la respuesta puede estar en que los responsables de las pruebas deben establecer desde su diseño estrategias que respondan a las necesidades de la evaluación como pueden ser: establecer la validez entre formas y años; garantizar la justicia y equidad de los resultados para los aspirantes

que apliquen cualquiera de las formas de la prueba; para la seguridad de administración de las pruebas en donde se aplique más de una ocasión a más de un grupo de sustentantes, lo que puede conducir a una sobreexposición de los reactivos por el número de observaciones, amenazando la seguridad de esta y, cada vez más, en los programas de evaluación se vislumbra la necesidad de la elaboración constante de reactivos como un proceso iterativo en el mantenimiento de la prueba.

Por lo que, el objetivo de este apartado es proporcionar una perspectiva de los métodos de equiparación y brindar un recurso útil para informar sobre las decisiones que se necesitan considerar al equiparar las formas de una prueba. Como tal, puede ser una guía para formular soluciones que sean prácticas, factibles y técnicamente sólidas cuando se toman decisiones que involucran distintas formas de prueba a partir de un mismo constructo.

El primer punto importante que debemos definir es, qué es una forma de la prueba.

Una forma está definida como una de las posibles múltiples pruebas de un mismo instrumento que mide el mismo constructo, frecuentemente estas formas suelen denominarse formas alternativas. Cuando dos formas de una prueba se han equiparado con éxito, se puede interpretar válidamente que una forma de la prueba funciona como si tuviera el mismo significado sustantivo en comparación con la puntuación equiparada de la otra forma de prueba.

Es importante considerar que, en el campo de la evaluación educativa, tanto en pruebas de alto impacto como en las pruebas a gran escala, los sustentantes responden a la selección de un conjunto de reactivos destinados a representar la totalidad del dominio de los conocimientos, habilidades o tareas posibles, con el propósito de brindar a estos una descripción precisa de lo que ellos saben y pueden hacer. Por lo que, una parte importante radica en que las formas estén construidas de acuerdo con los mismos contenidos especificados en su diseño y sean desarrollas de forma equivalente en la dificultad a lo largo de toda la prueba.

Los estándares *educativos y pruebas psicológicas* representan un marco de referencia internacional que establecen criterios para el desarrollo y la evaluación de pruebas y prácticas de desarrollo de pruebas y enmarcan las pautas para determinar la validez de las interpretaciones de los puntajes de las pruebas para los usos previstos de mismas. (AERA, APA, y NCME, 2018, p.1). Particularmente es importante en el Estándar 5.6, refiere que en los programas de evaluación las pruebas demuestren la estabilidad de sus escalas en la que reportan sus puntajes a lo largo del tiempo. Asimismo, los que refieren a la Vinculación de puntajes en los estándares 5.12 al 5.20 que requieren respaldos de evidencia de la comparabilidad de puntajes de una forma a otra y de un año a otro, así como de reunir datos adecuados y aplicar la metodología estadística apropiada para realizar la equiparación de puntajes en las formas alternativas de la prueba.

Un primer paso para la introducción a la equiparación es distinguir entre los conceptos que se han utilizado en la literatura; Holland y Dorans (2006) hicieron distinciones entre diferentes tipos de alineamiento (*linking*) y enfatizaron que estas distinciones están relacionadas con, cómo se utilizan e interpretan las puntuaciones vinculadas, dividiendo los métodos de alineamiento en tres categorías básicas: predicción, alineación de escala y

equiparación, considerando esencial comprender por qué estas categorías difieren y por qué pueden confundirse en la práctica.

El método de predicción para Dorans, Moses y Eignor (2010) tiene como objetivo predecir el puntaje de un sustentante en una prueba basándose en otra información sobre ese mismo sustentante. Por ejemplo, las puntuaciones de otras pruebas, calificaciones en cursos, el promedio de calificación, entre otros. Las relaciones establecidas mediante la predicción no son simétricas, por lo que no puede utilizarse para equiparar puntuaciones o producir puntuaciones con propiedades de puntuación comparables.

El método de alineación de escalas se refiere a las formas de pruebas que no se han diseñado de acuerdo con las mismas especificaciones; es decir, formas que pueden diferir en longitud o contenido; en este caso, las escalas vinculadas se consideran similares, pero no intercambiables. Esta relación se lleva a cabo a través de una función de *linking*. De acuerdo con Dorans et al. (2010) la alineación de escalas *(linking)* y la equiparación de puntuaciones a menudo se confunden porque los procedimientos estadísticos utilizados para la alineación de escalas también se pueden utilizar para equiparar pruebas; por lo que la alineación de escalas no puede ser considerada para equiparar puntuaciones.

En cambio, el método de equiparación se define en una relación estadística funcional entre múltiples distribuciones de puntaje de formas de una prueba y, por lo tanto, entre múltiples escalas de puntaje. Cuando las formas de las pruebas se han creado de acuerdo con las mismas especificaciones y son similares en características estadísticas esta relación funcional se denomina función de equiparación y sirve para convertir las puntuaciones de una escala directamente a sus valores equivalentes en otra.

A diferencia de *linkin*g, la equiparación apoya la afirmación de que un sustentante que aplica la prueba y obtiene una puntuación determinada en un año sabe y se comporta de manera similar a otro sustentante que obtiene el puntaje equivalente en el año siguiente. Esto es especialmente importante para mantener un significado equiparable de puntos de corte y niveles de rendimiento de uno año a otro, garantizando la equidad de los resultados entre las diferentes poblaciones que aplican cualquiera de las formas de la prueba.

PROPIEDADES QUE DEBE CUMPLIR PARA LA EQUIPARACIÓN

En el diseño se puede comprender que, en caso de ser requerido, se utilice la metodología de equiparación para corregir las diferencias en dificultad de las formas de la prueba. Lord (1980) establecieron requisitos específicos para equiparar puntajes entre formas y así definir las diferencias entre las distintas metodologías de alineamiento estas son: igualdad de constructos, confiabilidad similar, invarianza poblacional, simetría y equidad. De este modo, es necesario considerar (Dorans y Holland, 2000):

1) Igualdad de construcción: las distintas formas de la prueba deben ser medidas a partir del mismo constructo (aptitud, rasgo latente, habilidad).
2) Confiabilidad similar: las formas deben tener un nivel similar de confiabilidad.

3) Simetría: la transformación de equiparación para mapear las puntuaciones de B a los de A deberá ser el inverso de la transformación de equiparación para mapear las puntuaciones de A a las de B.
4) Equidad: deberá ser una cuestión de indiferencia para un examinado en cuanto a cuál de las formas de la prueba realiza realmente el examinado.
5) Invarianza de la población: la función de igualación utilizada para vincular las puntuaciones de A y B debe ser la misma, independientemente de la elección de la subpoblación de la que se deriva.

En la práctica, es difícil que se cumpla el conjunto de condiciones establecidas y no existe un acuerdo sobre cuáles son realmente las que se deben exigir en el proceso de equiparación. Sin embargo, con respecto a las mejores prácticas de la evaluación Dorans et al. (2010), consideran que los requisitos de igualdad de construcción y confiabilidad significan que las distintas formas de las pruebas deben construirse a partir de las mismas especificaciones, mientras que la simetría excluye los métodos de regresión como una forma de equiparación de prueba. Asimismo, en lo que respecta a la Equidad estos comprenden los dos primeros requisitos. Sin embargo, este requisito es difícil de evaluar empíricamente y su uso es principalmente teórico (Lord, 1980). Por último, Holland y Dorans (2006), consideraron que se pueden utilizar los dos primeros requisitos para explicar, si las dos pruebas miden cosas diferentes o no son igualmente confiables, por lo que, los métodos de vinculación estándar no producirán resultados que sean invariables en ciertas subpoblaciones de examinados.

DISEÑO PARA EL LEVANTAMIENTO DE DATOS

Para realizar la equiparación es preciso vincular la información de las formas, lo que implica recoger datos en muestras de sujetos. Por lo que en la literatura se han propuesto diversas formas de recolección, conocidas como diseños de equiparación (Kolen y Brennan, 2010; Holland y Dorans, 2006; y von Davier et al., 2004). Estos autores consideran que los diseños más utilizados en la práctica son: Diseño de un solo grupo, Diseño de grupo único contrabalanceado, Diseño de grupos equivalentes o grupos aleatorios y Diseño de grupos no equivalentes con reactivos comunes.

Diseño de un solo grupo

En los diseños de un solo grupo se administra en el mismo grupo de sujetos las dos formas de la prueba que se desean equiparar. Ambas formas deben medir las mismas especificaciones y presentar el mismo grado de dificultad. El diseño de un solo grupo puede proporcionar resultados de equiparación precisos con tamaños de muestra relativamente pequeños.

Figura 1. Diseño de un solo grupo

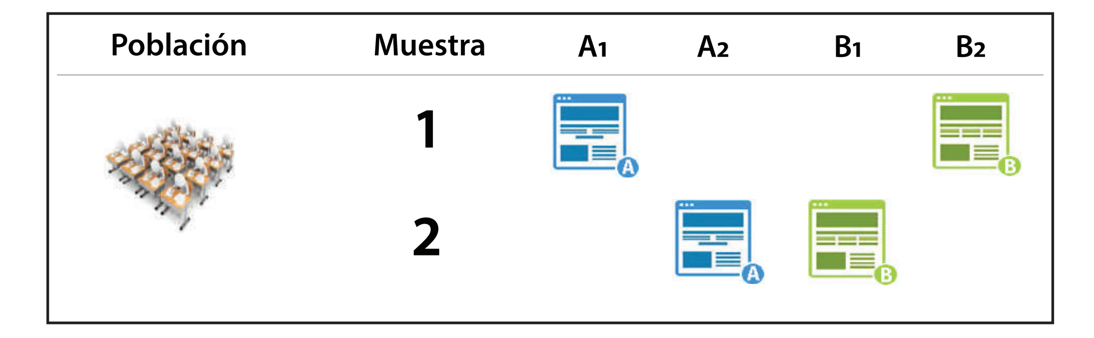
Población	Muestra	A	B
	1		

Este diseño se debe tener en cuenta que derivado del orden de administración de las formas de la prueba y los efectos de fatiga o aprendizaje que este puede producir, se debe asumir que el valor de las puntuaciones obtenidas por los sujetos en la segunda forma de la prueba, no están afectadas por habérseles administrado en una primera forma.

Diseño de un solo grupo contrabalanceado

Para asegurar la inexistencia de estos efectos, es recomendable utilizar una variante, el diseño de un solo grupo contrabalanceado, es una manera de poder evitar los posibles efectos del orden de administración de las dos formas de la prueba. En este caso, subdivide a los sujetos en dos grupos incluyendo en cada uno 50% de la muestra. A continuación, se administra a ambos subgrupos las dos formas de la prueba en orden inverso. De esta forma, se puede asegurar que ambas formas se verán afectadas por igual, por los efectos del orden de aplicación. En la figura 2 se observa que A_1 es cuando se toma primero y A_2 cuando se toma en segundo lugar, y de manera similar para B_1 y B_2, se describe el diseño de los grupos contrabalanceados.

Figura 2. Diseño de un solo grupo con contrabalanceo



Diseño de grupos equivalentes (grupos aleatorios)

En este diseño se extraen de la población de forma aleatoria dos muestras de sujetos, y a cada muestra se le aplica una forma de la prueba. Por lo tanto, cada sujeto responde solamente a una de las formas.

Figura 3. Diseño de grupos equivalentes

Población	Muestra	Forma A	Forma B
	1	A	
	2		B

En la mayoría de las aplicaciones para obtener muestras aleatorias y equivalentes alternar las formas en cada grupo, de tal manera, que al primer sujeto se le entregue la forma A, al segundo la forma B, y así sucesivamente. Este tipo de muestreo se le denomina muestra en espiral. El diseño es bastante conveniente de administrar, no requiere que las dos pruebas tengan reactivos en común, sin embargo, este diseño se puede utilizar incluso cuando tienen reactivos en común. Una limitación es que requiere grandes tamaños de muestra para tener resultados de equiparación precisos.

Diseño de grupos no equivalentes con reactivos comunes

En los diseños de prueba de anclaje hay dos poblaciones, P y Q, con una muestra de sustentantes de P, a quienes se les administra la forma de la prueba A, y una muestra de Q quien se les administra la prueba B. La diferencia con el diseño de grupos equivalentes consiste en que ambas muestras no tienen por qué ser equivalentes entre sí y, además, ambas muestras se les aplica un subconjunto de reactivos de prueba anclajes X en cada una de las formas de prueba deben equipararse. Por lo tanto, cada sujeto contesta un subconjunto de reactivos de la prueba diferente y uno común. von Davier et al. (2004). Kolen y Brennan (2010).

Figura 4. Diseño de grupos no equivalentes con reactivos comunes

Población		Muestra	A	X	B
	P	1	A	X	
	Q	2		X	B

La selección de reactivos destinados a cumplir la función de anclas es particularmente importante. La representación del contenido proporcional de los reactivos en el conjunto de ancla debe ser similar a la representación de contenido proporcional de toda la

forma de prueba, incluso hasta el punto de considerar que el conjunto de ancla es una "pequeña-versión" de la forma de prueba completa (Kolen y Brennan, 2010).

Este diseño presenta dos posibles modalidades: la prueba de anclaje interno y la prueba de anclaje externo.

La prueba de anclaje interno se utiliza un conjunto de reactivos comunes en ambas formas de la prueba y estos aparecen intercalados con el resto de los reactivos que son propios de las dos formas A y B, cuyas puntuaciones se quieren equiparar. Las puntuaciones obtenidas en los reactivos comunes se incluyen en la puntuación total de los sujetos en la prueba.

Figura 5. Reactivos de anclaje internos integrados

Forma A de la prueba	
Reactivo 1	A
Reactivo 2	Ancla
Reactivo 3	A
Reactivo 4	A
Reactivo 5	A
Reactivo 6	Ancla
Reactivo 7	A
Reactivo 8	Ancla
Reactivo 9	A
Reactivo 10	A
Reactivo 11	Ancla
Reactivo 12	A
Reactivo 13	A
Reactivo 14	Ancla
Reactivo 15	A
Reactivo 16	A
Reactivo 17	A
Reactivo 18	Ancla

Reactivos de anclaje internos integrados

Reactivo 2
Reactivo 6
Reactivo 8
Reactivo 11
Reactivo 14
Reactivo 18

Forma B de la prueba	
Reactivo 1	B
Reactivo 2	Ancla
Reactivo 3	B
Reactivo 4	B
Reactivo 5	B
Reactivo 6	Ancla
Reactivo 7	B
Reactivo 8	Ancla
Reactivo 9	B
Reactivo 10	B
Reactivo 11	Ancla
Reactivo 12	B
Reactivo 13	B
Reactivo 14	Ancla
Reactivo 15	B
Reactivo 16	B
Reactivo 17	B
Reactivo 18	Ancla

En la prueba de anclaje externo los ítems comunes aparecen formando una prueba independiente y las puntuaciones obtenidas por los sujetos en esa prueba no se utilizan en el cálculo de la puntuación total de los sujetos en las formas a equiparar.

Figura 6. Reactivos de anclaje externos integrados

	Forma A de la prueba	
Bloque 1 de la prueba	Reactivo 1	A
	Reactivo 2	A
	Reactivo 3	A
	Reactivo 4	A
	Reactivo 5	A
	Reactivo 6	A
	Reactivo 7	A
	Reactivo 8	A
	Reactivo 9	A
	Reactivo 10	A
	Reactivo 11	A
	Reactivo 12	A
Bloque 2	Reactivo 13	Ancla
	Reactivo 14	Ancla
	Reactivo 15	Ancla
	Reactivo 16	Ancla
	Reactivo 17	Ancla
	Reactivo 18	Ancla

	Forma B de la prueba	
Bloque 1 de la prueba	Reactivo 1	B
	Reactivo 2	B
	Reactivo 3	B
	Reactivo 4	B
	Reactivo 5	B
	Reactivo 6	B
	Reactivo 7	B
	Reactivo 8	B
	Reactivo 9	B
	Reactivo 10	B
	Reactivo 11	B
	Reactivo 12	B
Bloque 2	Reactivo 13	Ancla
	Reactivo 14	Ancla
	Reactivo 15	Ancla
	Reactivo 16	Ancla
	Reactivo 17	Ancla
	Reactivo 18	Ancla

Reactivos de anclaje internos integrados

En el primer caso se consideran reactivos de anclaje y en el segundo de prueba de anclaje. En ambos casos los ítems comunes deben ser lo más parecidos posible a los de las dos formas, aunque no sea condición imprescindible.

MÉTODOS DE EQUIPARACIÓN

Existen varias herramientas o procedimientos para la equiparación de las formas de la prueba, algunos asociados con la teoría clásica de los test (TCT) y otros con la teoría de respuesta al ítem (TRI). Todos estos procedimientos se pueden utilizar para equiparar y otros tipos de vinculación. Sin embargo, es esencial reconocer que cuando se usa para equiparación, estos procedimientos se aplican a las pruebas que han sido construidas para ser paralelos de modo que las puntuaciones en las formas tengan el mismo significado e interpretación.

Equiparación Identidad

La equiparación identidad es la más simple de todas, ya que las formas se consideran iguales y no es necesario equipararlas, considerando que un puntaje en la forma inicial (forma A) es equivalente a un puntaje idéntico en la forma nueva (forma B). La función matemática del método de identidad es:

$$B = ID\ B\ (A) = A$$

Donde, *A* se refiere a la puntuación bruta obtenida de la forma *A*, y *B* es la puntuación bruta equivalente a *A* para la forma B.

En la Gráfica 1, se puede observar la relación de identidad entre las dos formas en el puntaje 20. No es necesario un número mínimo de sustentantes, si el número de sustentantes es menor de 100 en alguna de las formas se recomienda utilizar el método de equiparación identidad, cuando las formas de prueba se consideran paralelas (Kolen y Brennan, 2010).

Equiparación de la media

Para este método de equiparación se considera que la forma A difiere de la B, esto debido a diferencias en dificultad. Por lo que, esta cantidad se considera constante a lo largo de la escala de puntuaciones, es decir, que, considerando un ejemplo, si A es más fácil en 4 puntos, estos 4 puntos son los mismos para los sujetos de alta habilidad que para los de baja. El supuesto básico es que las diferencias de los sujetos en las dos formas son iguales:

$$y - \mu\ (B) = x - \mu(A)$$

$y = m_B(x)$transformación a la forma B en función del puntaje de la forma A

$$m_B(x) = x - \mu(A) + \mu(B)$$

Donde:
x = un puntaje particular de la forma A
y = un puntaje particular de la forma B

$m_B(x)$ = un puntaje x en la forma A transformado a la escala de la forma B
$\mu(A)$ = media del puntaje de la forma A
$\mu(B)$ = media del puntaje de la forma B

Equiparación lineal

La equiparación lineal es una herramienta que se utiliza principalmente en TCT para determinar puntuaciones equivalentes entre dos formas de prueba paralelas que deben equipararse y se basa en la suposición de que las distribuciones de los puntajes de las formas A y la forma B son iguales, pero sus medias y desviaciones estándar son diferentes (Crocker y Algina, 1986).

Angoff (1984) definió la equiparación lineal como puntajes que son equivalentes cuando los puntajes en dos formas de prueba corresponden a las mismas desviaciones de puntaje estándar. Los puntajes en las formas con una distancia igual con respecto a sus medias (en unidades de desviación estándar) son considerados equivalentes.

$$\frac{x-\mu(A)}{\sigma(A)} = \frac{y-\mu(B)}{\sigma(B)}$$

$y = l_y(x)$ transformación a la forma B en función del puntaje de la forma A

$$l_Y(x) = \frac{\sigma(B)}{\sigma(A)}x + \left[\mu(B) - \frac{\sigma(B)}{\sigma(A)}\mu(A)\right]$$

donde:

$l_y(x)$ = un puntaje x en la forma A convertido a la escala de la forma B
$\mu(A)$ = media de la forma A
$\mu(B)$ = media de la forma B
$\sigma(A)$ = desviación estándar de la forma A
$\sigma(B)$ = desviación estándar de la forma B

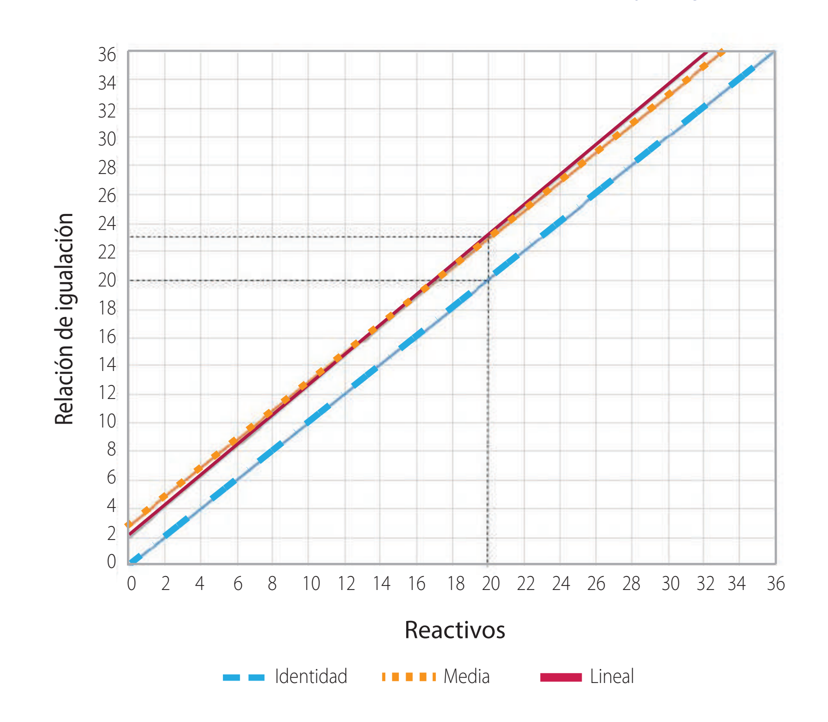
En la ecuación lineal toma su nombre del hecho de que la relación entre las puntuaciones de la forma de la prueba A y la forma de prueba B se puede mostrar como una línea recta en un gráfico. Por tanto, la línea representa la relación equivalente para todas las puntuaciones posibles. Como se muestra en la Gráfica 1, la forma de la prueba B en una prueba de 36 reactivos parece tener reactivos más difíciles que la Prueba A con 36 reactivos, basada en las puntuaciones medias (Prueba A media = 18.5 y Prueba B media = 17).

La Gráfica 1 muestra algunas de las limitaciones de este método, teniendo en cuenta que las puntuaciones de la forma B (media 20) son discretas: no tienen valores decimales. Pero no se transforman en puntajes brutos discretos en la forma A. Por ejemplo, un puntaje de 20 en la forma B es el equivalente ajustado del puntaje 18.5 en la forma A. Sin

embargo, los estudiantes que toman la forma A no pueden recibir un puntaje bruto de 18.5. Para abordar el problema de las puntuaciones equivalentes no discretas, se han utilizado varios enfoques para redondear las puntuaciones equivalentes de manera que les permitan informar puntuaciones discretas. Sin embargo, el redondeo introduce su propio tipo de error de equiparación.

Asimismo, se muestra una relación lineal en la que una puntuación muy alta en la forma B da como resultado un puntaje que está fuera del rango de puntajes posibles para la forma A. El gráfico parece sugerir que un puntaje de 36 en la forma B equivale a una puntuación de 39.9 o más en la forma A, lo cual no es posible. Sin embargo, Esto no es un error en la forma en que se traza el gráfico; si no, es parte de la naturaleza de la equiparación lineal.

Gráfica 1. Relación entre equiparación de identidad, de media y de igualación lineal



Método equipercentil

El método equipercentil proporciona precisión en la equiparación de resultados a lo largo de toda la escala de puntuación. También permite una mayor precisión que la ecuación lineal cuando las formas de prueba difieren en el nivel de dificultad general (Kolen y Brennan, 2010).

Para definir los rangos percentilares:

K_x = Número de reactivos en la forma A
f(x) = Proporción de sustentantes que tuvieron el puntaje *x*
f(x) ≥ para los puntajes enteros $x = 0,1, K_x$;
f(x) = 0 para otro puntaje

$$\sum f(x) = 1$$

f(x) = Proporción acumulada igual o debajo del puntaje

$$f(x) = 0 \; para \; x < 0; y$$

$$f(x) = 1 \; para \; x > K_x$$

$0 \leq F(x) \leq 1$ para x = 0,1, K_x;

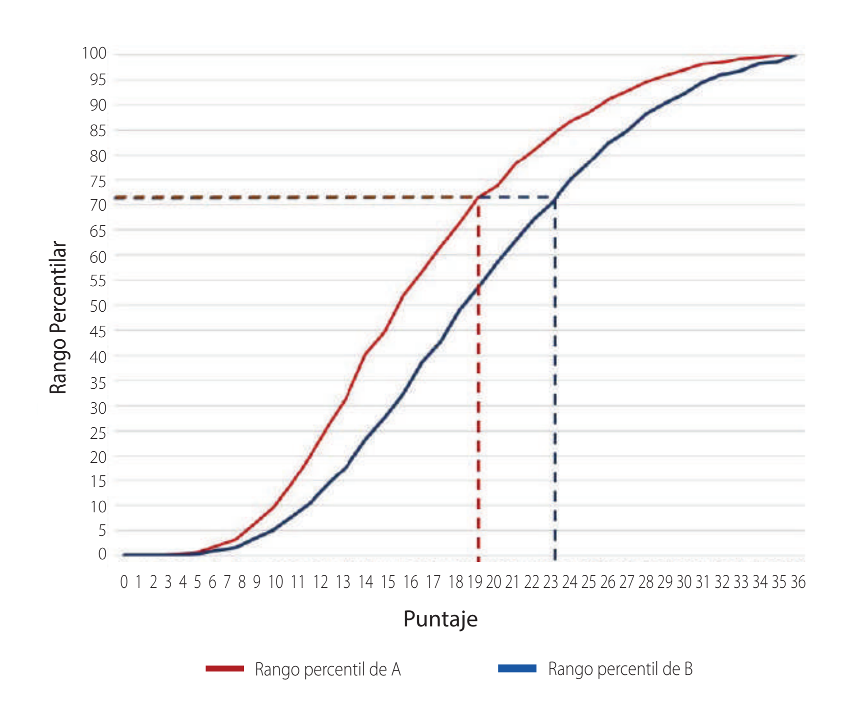
Si tenemos un valor de que no es entero, entonces x^* es el entero más cercano tal que $x^* - .5 \leq x < x^* + .5$

La función para calcular rango percentilar de la forma A es:

$$P(x) = 100\{F(x^* - 1) + [x - (x^* - .5)][F(x^*) - F(x^* - 1)]\},$$
$$-.5 \leq x < K_x + .5$$
$$= 0, x < -.5$$
$$= 100, x \geq K_x + .5$$

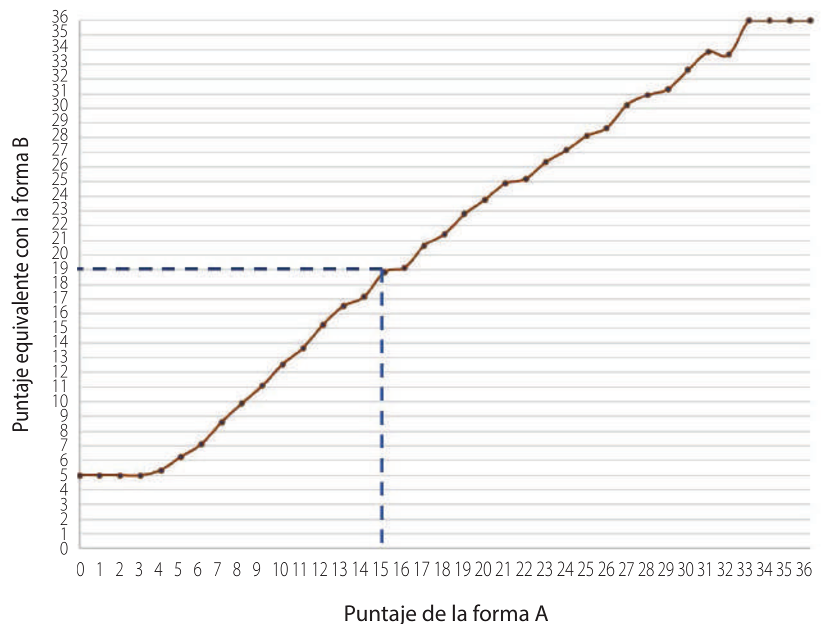
Se puede representar gráficamente las dos distribuciones de percentiles. Para ello, en el eje de abscisas ponemos las puntuaciones de la prueba A y de la forma B. En el eje de ordenadas los rangos percentiles. A continuación, trazan la curva correspondiente a cada prueba. Por lo que se obtiene las puntuaciones equivalentes en las dos formas A y B. En la Gráfica 2, se observa que el puntaje por debajo de 19 que representa 71% de los sustentantes para la forma A y por debajo del puntaje de 23 representa el mismo 71% en la forma B.





En la Gráfica 3 de equivalencias muestra las relaciones equipercentilares entre la forma A y la forma B de los datos, se observa que el puntaje 15 de la forma A es equivalente a 19 con la forma B.

Gráfica 3. Comparación entre los rangos de las dos formas A y B



DISEÑO DE GRUPOS NO EQUIVALENTES CON REACTIVOS COMUNES

Para estos métodos de equiparación, al ser bajo el diseño de grupos no equivalentes, consideran dos poblaciones diferentes: la población que contestó la forma A y la población que contestó la forma B; sin embargo, la función de equiparación está definida para una sola población.

En la literatura es uno de los diseños que más se utilizan. En cada muestra de sujetos se administra solamente una forma de la prueba, con la particularidad de que en ambas muestras se administra un conjunto de reactivos en común, que permite establecer la equivalencia entre las formas a equiparar.

Población sintética

El método de equiparación involucra dos poblaciones diferentes. Sin embargo, una función de equiparación de puntajes se considera como definida sobre una población única. Por lo tanto, las poblaciones 1 y 2 que corresponden a las poblaciones donde se aplicó la forma A (inicial) y B (nueva), deben ser combinadas para obtener una población única con el fin de definir una relación de equiparación.

Esta única población se conoce como población sintética Braun y Holland (1982), en la cual se le asignan pesos w_1 y w_2 a las poblaciones 1 y 2, respectivamente, esto es, $w_1 + w_2 = 1$ y $w_1, w_2 \geq 0$. Lo recomendable es utilizar

$$w_1 = \frac{N_1}{N_1 + N_2}$$

y

$$w_2 = \frac{N_2}{N_1 + N_2}$$

Donde N_1 corresponde al tamaño de la población 1 y N_2 corresponde al tamaño de la población 2.
Los puntajes de la forma nueva, aplicada a la población 1, serán denotados por X; Los puntajes de la forma antigua, aplicada a la población 2, serán denotados por .

Anclaje interno

Los puntajes comunes serán denotados por V y expresamos que los reactivos comunes corresponden a un anclaje interno cuando V se utiliza para calcular los puntajes totales de ambas poblaciones.

Métodos Lineales

Usando el concepto de población sintética, la relación lineal de equiparación de puntajes para el diseño de grupos no equivalentes con reactivos comunes se escribe de la siguiente forma:

$$l_{Y_s}(x) = \frac{\sigma_s(Y)}{\sigma_s(X)}[x - \mu_s(X)] + \mu_s(Y)$$

Donde denota la población sintética y

$$\mu_s(X) = \mu_1(X) - w_2\gamma_1[\mu_1(V) - \mu_2(V)]$$

$$\mu_s(Y) = \mu_2(Y) + w_1\gamma_2[\mu_1(V) - \mu_2(V)]$$

$$\sigma_s^2(X) = \sigma_1^2(X) - w_2\gamma_1^2[\sigma_1^2(V) - \sigma_2^2(V)] + w_1 w_2\gamma_1^2[\mu_1(V) - \mu_2(V)]^2$$

$$\sigma_s^2(Y) = \sigma_2^2(Y) + w_1\gamma_2^2[\sigma_1^2(V) - \sigma_2^2(V)] + w_1 w_2\gamma_2^2[\mu_1(V) - \mu_2(V)]^2$$

Donde los subíndices 1 y 2 se refieren a las poblaciones 1 y 2 respectivamente.

$$\gamma_1 = \frac{\sigma_1(X,V)}{\sigma_1^2(V)}$$

y

$$\gamma_2 = \frac{\sigma_2(X,V)}{\sigma_2^2(V)}$$

En la variante del método que se va a utilizar, las *y' s* se pueden expresar de la siguiente manera:

$$\gamma_1 = \frac{\sigma_1^2(X)}{\sigma_1(X,V)}$$

$$\gamma_2 = \frac{\sigma_2^2(Y)}{\sigma_2(Y,V)}$$

En la aplicación de este método basta con reemplazar estos coeficientes en las ecuaciones lineales antes descritas. Kolen y Brennan en 2010, proporcionan justificaciones para usar esta aproximación.

MÉTODOS DE EQUIPARACIÓN BASADOS EN EL MODELO DE RESPUESTA AL ÍTEM (TRI)

Existen una variedad de enfoques para equiparar en TRI, entre los que se encuentran los más utilizados que son la equiparación a través de reactivos comunes. Se calibran cada una de las formas de la prueba usando una métrica común, completamente por separado. Posteriormente se evalúa la relación entre los parámetros TRI en cada forma y se usa para estimar la relación para convertir las puntuaciones de los sustentantes. Lo que se hace es alinear los parámetros TRI de los reactivos comunes (ancla) se pueda aplicar esa transformación lineal a las puntuaciones. Los métodos de calibración se pueden analizar bajo los métodos de momento (media-media, media-sigma) y métodos de curva característica (Stocking Lord y Haebara). Las dos escalas del mismo constructo difieren solo en una transformación lineal simple, por lo que requiere encontrar la pendiente y la intersección de esa transformación.

Métodos de transformación Media-Sigma y Media-Media

Considerando que I y J son dos escalas construidas como escalas de TRI de 3 parámetros, que toma valores de θ para las dos escalas se relacionan de la siguiente forma:

$$\theta_{Ji} = A\theta_{Ii} + B$$

donde A y B son constantes en la ecuación lineal y θ_{Ji} y θ_{Ii} son los valores de θ para el individuo i en las escalas J e I. Los parámetros de los reactivos en las dos escalas están dados por:

$$a_{Jj} = \frac{a_{Ij}}{A},$$
$$b_{Jj} = Ab_{Ij} + B,$$
$$c_{Jj} = c_{Ij}$$

donde y a_{Jj}, b_{Jj} y c_{Jj} son los parámetros del ítem j de la escala J, a_{Ij}, b_{Ij} y c_{Ij} son los parámetros del ítem j de la escala I.

Para dos individuos i e i^* y dos ítems j y j^*, A y B pueden ser expresados como:

$$A = \frac{\theta_{Ji} - \theta_{Ji^*}}{\theta_{Ii} - \theta_{Ii^*}} = \frac{b_{Jj} - b_{Jj^*}}{b_{Ij} - b_{Ij^*}} = \frac{a_{Ij}}{a_{Jj}}$$
$$B = b_{Jj} - Ab_{Ij} = \theta_{Ji} - A\theta_{Ii}$$

De las anteriores ecuaciones se deriva que A y B pueden ser expresados como:

$$A = \frac{\sigma(b_J)}{\sigma(b_J)}$$
$$= \frac{\mu(a_I)}{\mu(a_J)}$$
$$= \frac{\sigma(\theta_J)}{\sigma(\theta_I)},$$
$$B = \mu(b_J) - A\mu(b_I)$$
$$= \mu(\theta_J) - A\mu(\theta_I)$$

donde las medias $\mu(b_J)$, $\mu(b_I)$, $\mu(a_J)$, $\mu(a_I)$ y las desviaciones estándar $\sigma(b_J)$ y $\sigma(b_I)$ son definidas sobre los parámetros de los ítems en cada escala *I* y *J*. Las medias $\mu(\theta_J)$ y $\mu(\theta_I)$ se definen sobre los sustentantes, de acuerdo con los parámetros relacionados con la habilidad en cada escala.

Método de transformación Media-Sigma

Es un proceso de conversión de puntaje definido por Marco (1977). En este método, la desviación estándar se utiliza en la determinación de la curva de la ecuación; y las dificultades promedio de las pruebas se utilizan en la determinación de la constante de ecuación.

$$A = \frac{\sigma(b_J)}{\sigma(b_I)}$$

$$B = \mu(b_J) - A\,\mu(b_I)$$

$\mu(b_I)$: media de los parámetros de dificultad para la escala *I*
$\mu(b_J)$: media de los parámetros de dificultad para la escala *J*
$\sigma(b_I)$: desviación estándar de los parámetros de dificultad para la escala *I*
$\sigma(b_J)$: desviación estándar de los parámetros de dificultad para la escala *J*
A: pendiente
B: intercepto

Método de transformación media-media

El método media-media definido por Loyd y Hoover (1980) calcula los coeficientes A y B utilizando las medias de los parámetros de discriminación y de dificultad. La media de los parámetros de los reactivos comunes para despejar el valor de la constante A y la media de los parámetros de los reactivos comunes para despejar el valor de la constante B, esto es:

$$A = \frac{\mu(a_I)}{\mu(a_J)}$$

$$B = \mu(b_J) - A\mu(b_I)$$

$\mu(a_I)$: media de los parámetros de discriminación para la escala *I*
$A\mu(b_I)$: media de los parámetros de dificultad para la escala *I*
$\mu(a_J)$: media de los parámetros de discriminación para la escala *J*
$\mu(b_J)$: media de los parámetros de dificultad para la escala *J*
A: Pendiente
B: intercepto

Método Haebara

En el método Haebara, la curva de la ecuación y la constante de la ecuación se obtienen utilizando la diferencia entre las curvas características del reactivo. Este enfoque es desarrollado por Haebara (1980). Para los sustentantes que tienen un cierto nivel de habilidad, la diferencia entre las curvas características del reactivo es la suma de los cuadrados de las curvas características del reactivo que pertenecen a cada reactivo. Se intenta encontrar la constante de ecuación y la curva de ecuación que minimizan esta diferencia (Kolen y Brennan, 2004). Para un θ_i dado, la suma, sobre los reactivos, del cuadrado de las diferencias se puede ver como:

$$Hdiff(\theta_i) = \sum_{j:V}\left[p_{ij}\left(\theta_{Ji}, \hat{a}_{Jj}, \hat{b}_{Jj}, \hat{c}_{Jj}\right) - p_{ij}\left(\theta_{Ji}, \frac{\hat{a}_{Ij}}{A}, A\hat{b}_{Ij} + B, \hat{c}_{Ij}\right)\right]^2$$

La sumatoria es sobre los reactivos comunes *(j:V)*. En esta ecuación, la diferencia entre la curva característica de cada reactivo en las dos escalas se eleva al cuadrado y se suma. *Hdiff* es entonces acumulada sobre los sustentantes. El procedimiento de estimación encuentra A y B que minimizan el siguiente criterio:

$$Hcrit = \sum_{i} Hdiff(\theta_i)$$

Método Stocking-Lord

En el método Stocking-Lord la curva de la ecuación y la constante de la ecuación se obtienen utilizando la diferencia entre las curvas características del ítem. A diferencia del enfoque de Haebara, en Stocking-Lord (1983), para los participantes que tienen un cierto nivel de habilidad, la diferencia entre las curvas características del ítem es el cuadrado de la suma de la diferencia entre las curvas características del ítem pertenecientes a cada ítem. Se intenta encontrar la constante de ecuación y la curva de ecuación que minimizan esta diferencia (Kolen y Brennan, 2004).

El enfoque de Stocking-Lord utiliza el cuadrado de la diferencia de las sumas sobre los reactivos, y la notación estadística se define como:

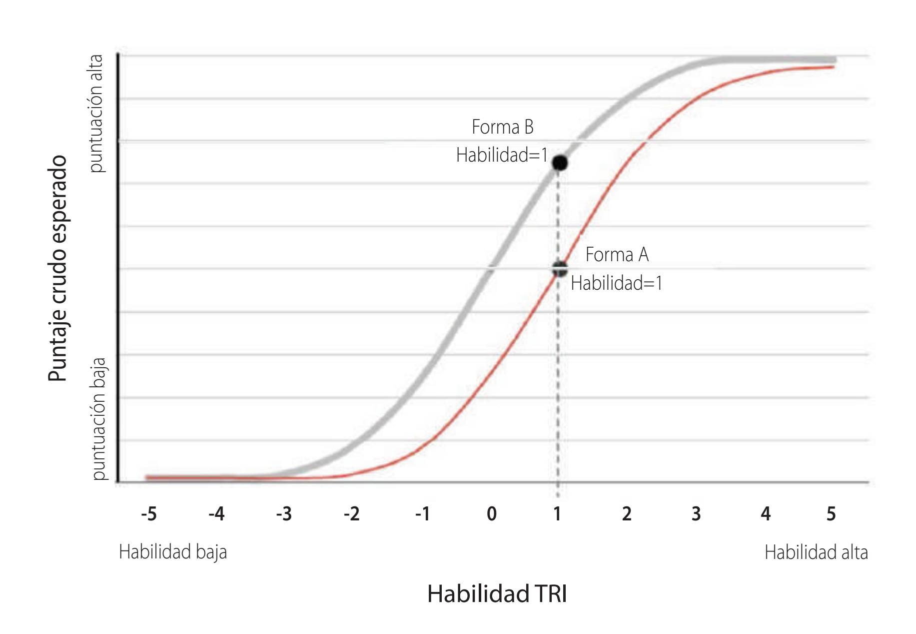
$$SLdiff(\theta_i) = \left[\sum_{j:V} p_{ij}\left(\theta_{Ji}, \hat{a}_{Jj}, \hat{b}_{Jj}, \hat{c}_{Jj}\right) - \sum_{j:V} p_{ij}\left(\theta_{Ji}, \frac{\hat{a}_{Ij}}{A}, A\hat{b}_{Ij} + B, \hat{c}_{Ij}\right)\right]^2$$

La expresión *SLdiff*(θ_i) es el cuadrado de las diferencias entre las curvas características de la prueba para una habilidad dada θ_i. El procedimiento de estimación es encontrar la combinación de A y B tal que se minimice el siguiente criterio:

$$SLcrit = \sum_{i} SLdiff\left(\theta_i\right)$$

A manera de ejemplo que se muestra en la gráfica 4, donde se observan las dos formas de prueba comparten reactivos comunes, pero cada forma también tiene su propio conjunto de reactivos únicos. En la forma de prueba A, el examen de la derecha se muestra más difícil ya que la misma posición de habilidad de TRI en el eje horizontal se asigna a una puntuación bruta más baja que la puntuación bruta de la forma de prueba B que corresponde a la misma habilidad (Ryan, y Brockmann, 2022).

Gráfica 4. Curvas de características de la prueba



La equiparación de TRI es compleja, tanto conceptual como procedimentalmente. La definición de TRI en puntajes equiparados se basa en una abstracción, más que en estadísticas que realmente se pueden calcular. TRI se basa en suposiciones sólidas que, a menudo, no son una buena aproximación a la realidad de las pruebas (Livingston, 2014).

En la actualidad existe una variedad de programas tanto gratuitos como de licencia, entre ellos el desarrollado por Kollen y Brenan, el paquete se llama *CIPE,* sirve para calcular varios procedimientos de equiparación (Kolen y Brennan, 2010), Asimismo otros paquetes del software estadístico R; el paquete *equate* (Albano, 2016) que reproduce los resultados del *CIPE.* Asimismo, los programas de TRI como BILOG-MG, ICL, MULTILOG, PARSCALE, Xcalibre, IRTEQ, entre otros.

Kolen y Brennan, (2010) recomiendan que para que cualquiera de estos métodos se utilice adecuadamente las especificaciones de la prueba, los datos recopilados y los procedimientos de estandarización y control de calidad deben ser adecuados. Sin embargo, para decidir qué métodos estadísticos implementar para una equiparación particular dependerá de las características de las situaciones de equiparación para las cuales cada uno de los métodos puede ser apropiado. Por lo debe considerar la literatura de investigación sobre métodos de equiparación y la realización de investigaciones para el programa de prueba en cuestiones prácticas en la equiparación en el que se va a hacer la equiparación. En la Tabla 1 se presenta una lista de características de las situaciones de equiparación basada en (Kolen y Brennan, 2010; p.305-310)

Tabla 1. Métodos de equiparación

Diseño	Método						
	Características	Identidad	Media	Lineal	Equipercentil	Rasch	TRI 3 PL
Grupos aleatorios equivalentes y Grupo No equivalente con reactivos en común	Control de calidad o condiciones de estandarización	Deficiente	Adecuada	Adecuada	Adecuada	Adecuada	Adecuada
	Muestras	Muy pequeñas o ningún dato	Muy pequeñas	Pequeñas	Grandes	Pequeñas	Grandes
	Forma de prueba similares	Dificultades de forma de prueba similares	Dificultades de forma de prueba similares	Dificultades de forma de prueba similares	Las formas de prueba pueden diferir en el nivel de dificultad más que para un método lineal	Dificultades de forma de prueba similares	Las formas de prueba pueden diferir en el nivel de dificultad más que para un método lineal
	Entendimiento de tablas de conversión o ecuaciones, en la realización de análisis	Sí	Sí	Sí	No	No	No
	Facilidad para explicar el procedimiento a personas que no son psicómetras	Sí	Sí	Sí	No	No	No
	Precisión de los resultados	No (tolera resultados inexactos)	Sí (alrededor de la media)	Sí (alrededor de la media)	Sí (a lo largo de la escala de puntuación)	Sí (alrededor de la media).	Sí (a lo largo de la escala de puntuación)
	Cumple los supuestos del modelo IRT	No aplica	No aplica	No aplica	No aplica	Sí	Sí
Grupos No equivalentes con reactivos en común No	Conjunto de reactivos en común que sean representativos del constructo	No	Sí	Sí	Sí	Sí	Sí
	Grupos de sustentantes con nivel de logro similar	No	Sí	Sí	Sí	Sí	Sí

A MANERA DE CIERRE

La equiparación juega un papel importante en el sistema de responsabilidad escolar y, en general, para las instancias evaluadoras ya que permite comparar los puntajes de una forma de prueba a otra, de una aplicación a otra o de un año a otro. Por lo tanto, la calidad técnica de los métodos de equiparación utilizados y la documentación de los procesos de equiparación empleados, son relevantes para cualquier sistema de rendición de cuentas destinado a reflejar el crecimiento anual y la mejora continua en los procesos de evaluación.

Asimismo, como se considera importante que las decisiones que se tomen al momento de equiparar sean mejor informadas mediante la descripción de problemas comunes relacionados con la equiparación y garanticen la equidad de los resultados para los sustentantes que contestaron cualquiera de las formas de la prueba.

El primer paso para respaldar una inferencia de equivalencia comienza con la construcción de las formas de prueba. Para ser equiparadas, las formas de prueba deben construirse de acuerdo con los mismos contenidos y especificaciones. Los elaboradores de las pruebas deben solicitar el modelo y las especificaciones de esta y cualquier otra documentación que respalde la inferencia de que las formas son equivalentes. La evidencia de los procedimientos psicométricos debe incluir datos que verifiquen que todas las suposiciones hechas para respaldar la equiparación han sido examinadas e informadas. En la equiparación de reactivos comunes, se debe presentar la evidencia de la estabilidad de los reactivos de anclaje.

Por lo anterior, de acuerdo con los estándares, la equiparación apoya la interpretación de los resultados y el uso para el cual fue diseñado el instrumento, abonando a las evidencias de validez que requiere el diseño y mantenimiento de la prueba.

Error de Equiparación

En los procedimientos de equiparación existen diferencias en las características de las muestras seleccionadas al azar lo que puede reflejar un error de muestreo. A esto se suman los propios procedimientos de equiparación que conllevan cierta imprecisión o error en el proceso de equiparación, así como las formas de la prueba tienen un cierto grado de error de medición.

Para Ryan y Brockmann, (2022) el concepto de error es la diferencia entre el valor observado o esperado y el valor real de algo, encontrado en el proceso de resolución del problema. Por lo que se debe describir el error como el nivel de precisión o certeza que se sabe que producen las puntuaciones de las pruebas, los muestreos, las estimaciones de parámetros o los procedimientos de equiparación.

Los estándares de la AERA, APA, y NCME, (2018) establecen en su estándar 5.13 que se deben estimar y reportar siempre que sea posible los errores estándares de las funciones de equiparación. Asimismo, consideran que debe presentarse en unidades de la escala de puntajes reportada. Así como para los programas de evaluación con puntajes de corte, el error de equiparación cercano al punto de corte es de primordial importancia. En los programas proporcionan una aproximación al error de equiparación, de acuerdo con el tipo de métodos de equiparación con que se esté haciendo el procedimiento.

REFERENCIAS

AERA, APA, & NCME (2014). *Standards for Educational and Psychological Testing.* American Educational Research Association, American Psychological Association, & National Council of Measurement in Education.

Albano, Anthony. (2016). equate: An R Package for Observed-Score Linking and Equating. *Journal of Statistical Software. 74*(8). https://doi.org/10.18637/jss.v074.i08

Angoff, W. H. (1984). Scales, norms, and equivalent scores. Princeton, NJ: Educational Testing Service.

Braun, H. I., & Holland, P. W. (1982). Observed-score test equating: A mathematical analysis of some ETS equating procedures. In P. W. Holland & D. B. Rubin (Eds.), Test equating (pp. 9–49). New York: Academic.

Brennan, R.J. (2006). Chained Linear Equating. CASMA Technical Note. Center for Advanced Studies in Measurement and Assessment, Universidad de Iowa, IA. Recuperado el 12 de Julio de 2010, de http://www.education.uiowa.edu/casma/documents/clinearreport3.pdf

Dorans, Neil & Moses, Tim & Eignor, Daniel. (2010). Principles and Practices of Test Score Equating. ETS Research Report Series. 2010. i-41. https://doi.org/10.1002/j.2333-8504.2010.tb02236.x

Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research. 22*, pp. 144-149.

Holland P, Dorans N (2006). Linking and Equating. In R Brennan (ed.), Educational Measurement, 4th edition, pp. 187–220. Greenwood, Westport.

Kolen, M. J., & Brennan, R. L. (2010). Test Equating, Scaling, and Linking: Methods and Practices (Second ed.). (Springer, Ed.) New York

Marco, G. L. (1977). Item Characteristic Curve Solutions to Three Intractable Testing Problems. *Journal of Educational Measurement, 14*(2), 139-160.

Loyd, B. H. and Hoover, H. D. (1980). Vertical Equating Using the Rasch Model. *Journal of Educational Measurement, 17*(3), 179-193.

Livingston, S. (2004). Equating Test Scores (Withot IRT). Educational Testing Service, Princeton, NJ.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Routledge.

Ryan, J. & Brockmann, F. (2009). *A practitioner's introduction to equating with primers on classical test and item response theory.* Washington, DC: Council of Chief State School Officer.

Stocking, M. & Lord, F. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement. 56*(4), pp. 570 – 584

von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). The kernel method of test equating. New York, NY: Springer.

Sección III.

INSTRUMENTOS Y ESTRATEGIAS DE EVALUACIÓN (ENFOQUE CUALITATIVO)

Capítulo 21

PORTAFOLIO

Ana Itzel Pascual Vigil, Cinthya Trejo Rojas

"La naturaleza de un portafolio es que siempre es un trabajo en progreso."
CAT SETO Y MEG ILASCO

INTRODUCCIÓN

En el proceso de enseñanza existen diversos factores que el profesor-facilitador debe armonizar para realizar su clase, entre estos se encuentran los contenidos curriculares, las estrategias de enseñanza y las formas de evaluación que, en conjunto, representan un desafío en el espacio educativo –aula, laboratorio, talleres, salas–. Una forma de afrontar este desafío reside en implementar estrategias que muestren otros aspectos relevantes que se viven dentro del proceso de formación y que impactan directamente en el contenido curricular, así como en los procesos de enseñanza-aprendizaje.

Aunado a este desafío, en la actualidad, los profesores-facilitadores también se enfrentan a fenómenos emergentes que impulsan a realizar ajustes severos para adaptar su enseñanza de acuerdo con el acceso a herramientas o medios que implican el uso de la creatividad y que probablemente se consolidaran como parte común del contenido curricular.

El profesor-facilitador puede superar estos desafíos con el uso del portafolio de evidencias, el cual es un instrumento de evaluación que permite acompañar el proceso de aprendizaje y dar respuesta a preguntas como: ¿qué ha aprendido mi alumno?, ¿qué información sobre su aprendizaje puedo observar en las evidencias del portafolio?, ¿cuáles son los temas de mayor interés para mis alumnos?, ¿qué medidas puedo implementar para ayudarlos en aquellos aprendizajes que no están logrando?, ¿cómo puedo involucrarlos para que le den sentido a sus aprendizajes?, ¿cómo puedo modificar mis estrategias de enseñanza de acuerdo a las necesidades de mis alumnos?, ¿qué elementos de mi proceso de enseñanza están dando mejores resultados y cuáles puedo mejorar?, ¿qué medios o herramientas son más prácticos para mis alumnos?

El portafolio permite dar testimonio del desarrollo, los avances y la creatividad que poseen los alumnos, facilitando la valoración de sus dificultades, necesidades y potencialidades.

Además, brinda al profesor-facilitador un acercamiento hacia sus alumnos, sus procesos y sus productos de aprendizaje de forma individual y grupal, generando nuevas oportunidades de mejora del proceso de enseñanza.

El propósito de este capítulo es orientar a los profesores en el uso del portafolio de evidencias, de forma presencial, mixta o a distancia, como herramienta que los ayude en el diseño e implementación de la evaluación del, para y como aprendizaje.

El capítulo está organizado en siete secciones a modo de preguntas. En la primera se explica qué es un portafolio de evidencias, en la segunda se detallan los pasos para diseñarlo, en la tercera se recomienda cómo aplicarlo, la cuarta consiste en describir el uso del portafolio en la educación a distancia, en la quinta se presentan ejemplos para ilustrar su uso, en la sexta se incluyen los recursos en línea y medios electrónicos para implementarlos de manera digital, por último en la séptima se presentan las conclusiones y reflexiones finales.

Este capítulo intenta ser un recurso de consulta flexible porque puede leerse acorde a sus necesidades. Se espera que contribuya con el desafío que representa la labor del profesor-facilitador y fomente la reflexión sobre el complejo proceso de enseñanza-aprendizaje.

¿QUÉ ES?

El portafolio es un instrumento en el que se integran o compendian trabajos realizados por el estudiante durante el ciclo de formación. Este le permite demostrar, por medio de la selección y organización de evidencias, los aspectos más significativos de su proceso de aprendizaje. Esta colección permite que tanto alumno como profesor-facilitador reconozcan, valoren y reflexionen sobre el crecimiento gradual de los aprendizajes logrados en la formación (Díaz-Barriga, Martínez-Romero y Heredia-Sánchez, 2012).

El concepto surgió inicialmente en otros ámbitos, como las artes y la fotografía, en los que se requiere de la presentación de evidencias para valorar los logros que se exponen a través de un conjunto de trabajos representativos (Gutiérrez-Morales y Sánchez-Mendiola, 2020).

El portafolio puede tomar muchos sentidos y formas diferentes dependiendo de la finalidad, el análisis y los objetivos que se le otorgue (Barragán-Gregori, 2009). Por ejemplo, al considerar una evaluación sumativa para una asignatura, el profesor-facilitador concentrará su atención en analizar las evidencias que integran al portafolio para conocer los resultados de su aprendizaje en los distintos temas que se abordaron a lo largo del proceso de enseñanza.

Las evidencias que el alumno puede incluir en un portafolio son diversas, entre estas están: ensayos, estudio de casos, anteproyectos de investigación, resolución de problemas, reportes, reseñas, cuadros comparativos, sinópticos, mentales, resúmenes, análisis de artículos académicos, diarios de campo, bitácoras, videos, presentaciones, audios, comentarios reflexivos, relatos, proyectos, trabajos colectivos, notas de campo y autoevaluaciones, entre otros (Colén, Giné e Imbernón, 2006, citado por Díaz-Barriga et al., 2012).

Existen tres tipos de portafolio:

- **Estructurado:** el profesor determina previamente las evidencias que debe presentar el portafolio, ya sea de manera individual o en equipos.
- **Semiestructurado:** el profesor establece algunas de las evidencias obligatorias que debe contener el portafolio y además brinda al alumno la posibilidad de agregar evidencias optativas que el aprendiz considere relevantes.
- **Libre:** el profesor brinda a sus alumnos la oportunidad de decidir las evidencias que entregarán en el portafolio.

El formato que se puede utilizar para integrar un portafolio puede ser físico o en línea. Por ejemplo, en un portafolio en línea se pueden incluir evidencias con diversos formatos como audio, video, gráficos, imágenes o fotografías, páginas web, entre otros; ello permite que los alumnos desarrollen habilidades tecnológicas. En la siguiente sección se explicarán las consideraciones para implementar el portafolio en una evaluación del, para y como aprendizaje.

Para utilizar el portafolio como evaluación es necesario conjugarlo con otros instrumentos como rúbricas, escalas, listas de cotejo o exposiciones (Barberá-Gregori, 2005, 2009).

El uso del portafolio requiere de una planeación adecuada, de otra forma se puede tornar en una sobrecarga de trabajo que impactará en el proceso educativo y en la falta de retroalimentación para los alumnos.

¿CÓMO LO DISEÑO?

Para implementar el portafolio como un instrumento de evaluación se sugiere analizar el programa de su asignatura para establecer los aprendizajes que evaluará, después se recomienda seguir el siguiente proceso:

Planeación

1) **Seleccione los aprendizajes esperados que evaluará.**
   Considere objetivos, contenidos, actividades y aprendizajes esperados, así como las características de sus alumnos y el medio en el que se desarrollará el portafolio: físico-presencial o tecnológico a distancia.

   Consejo 1. El uso del portafolio potencia la confianza, la independencia y la autonomía de los alumnos.

2) **Determine el tipo de evaluación que desea realizar con el portafolio.**
   La evaluación puede ser diagnóstica, formativa o sumativa.
3) **Seleccione el tipo de portafolio que se adecue al objetivo de su evaluación.**
   En caso de que el portafolio que desee implementar sea estructurado o semiestructurado,

puede seleccionar evidencias obligatorias u optativas. Para el portafolio libre es importante que el alumno decida qué entregará como evidencia de su aprendizaje.

Consejo 2. Resuelve el siguiente crucigrama: https://es.educaplay.com/recursos-educativos/10575112-portafolio.html

4) **Defina si el portafolio será elaborado de manera individual o como un trabajo en equipo.**
   Esto dependerá de las finalidades, tiempos y medios que se establezcan para la implementación de este instrumento.
5) **Establezca las actividades y el tiempo de integración del portafolio.**
   Considere las actividades en clase y extra-clase que pueden servir como evidencia y el tiempo necesario para hacer el análisis, la selección y organización de las evidencias.

   Consejo 3. El tiempo que puede considerar es: la elaboración de las evidencias, la retroalimentación al alumno y la revisión final del portafolio.

6) **Establezca quiénes participarán en la evaluación y de qué forma.**
   Puede elegir entre tres formas: la primera, en la que se valoran las evidencias del alumno directamente (heteroevaluación); la segunda, en la cual un alumno valora el trabajo de otro alumno (coevaluación); y la tercera, en la que el propio alumno valora su propio trabajo con la ayuda del docente (autoevaluación).
7) **Determine cuáles serán los criterios de evaluación y niveles de logro.**
   Es importante que defina los criterios que evaluará en el portafolio de manera general y, si lo considera necesario, para cada una de las evidencias.
8) **Seleccione un instrumento que acompañe la evaluación del portafolio.**
   Es importante que acompañe el portafolio con otro instrumento de evaluación que contenga los criterios anteriores.

   Consejo 4. Para acompañar la evaluación del portafolio puede elegir una lista de cotejo, rúbrica o escala.

Diseño

Para la implementación del portafolio en la evaluación se sugiere utilizar la siguiente **guía de instrucciones**:

1) **Incluya datos de identificación.**
   Nombre de la institución, nombre de la asignatura, objetivo del portafolio e instrucciones.
2) **Especifique qué tipo de portafolio debe desarrollar el alumno.**
   Puede ser estructurado, semiestructurado o libre.
3) **Determine el formato del portafolio.**
   Puede ser físico o en línea.

Para un portafolio estructurado, ver Ejemplo 1.
Incluya los pasos del 1 al 3 de la guía de instrucciones.

4) **Describa las secciones que integrarán al portafolio.**
   Las categorías en las que se puede organizar el portafolio pueden ser unidades temáticas, según el programa de estudios o tipo de actividades o clases, entre otras categorías.
5) **Defina las evidencias de cada sección.**
   Las evidencias tienen que facilitar tanto la observación del proceso de aprendizaje, como el proceso de creación del portafolio del alumno.
6) **Establezca las sesiones de retroalimentación que se realizarán durante el desarrollo del portafolio.**
   En estas sesiones el alumno y el profesor establecen un diálogo sobre los avances del portafolio y la construcción de su aprendizaje.
7) **Determine la fecha de entrega y la retroalimentación final.**
   Es importante considerar el tiempo suficiente para la revisión de los portafolios y la retroalimentación final para el alumno.

Para un portafolio semiestructurado, ver Ejemplo 2.
Incluya los pasos del 1 al 3 de la guía de instrucciones.

4) **Defina las evidencias obligatorias que debe contener el portafolio.**
   Seleccione las actividades que faciliten tanto la observación del proceso del alumno durante su aprendizaje, como el proceso de creación del portafolio.
5) **Señale la posibilidad de incluir evidencias optativas.**
   Este tipo de evidencias le permitirán al alumno demostrar sus valoraciones sobre su propio trabajo.
6) **Establezca las sesiones de retroalimentación que se realizarán durante el desarrollo del portafolio.**
   En estas sesiones el alumno y el profesor establecen un diálogo sobre los avances del portafolio y la construcción de su aprendizaje.
7) **Determine la fecha de entrega y la retroalimentación final.**
   Es importante considerar el tiempo suficiente para la revisión de los portafolios y la retroalimentación final para el alumno.

Para un portafolio libre, ver Ejemplo 3.
Incluya los pasos del 1 al 3 de la guía de instrucciones.

4) **Indique que el alumno es libre de elegir las evidencias que integrarán al portafolio.**
   Este tipo de evidencias busca fomentar que el alumno valore y seleccione su propio trabajo para representar el avance de su aprendizaje.
5) **Establezca las sesiones de retroalimentación que se realizarán durante el desarrollo del portafolio.**
   En estas sesiones el alumno y el profesor establecen un diálogo sobre los avances del portafolio y la construcción de su aprendizaje.
6) **Determine la fecha de entrega y la retroalimentación final.**
   Es importante considerar el tiempo suficiente para la revisión de los portafolios y la retroalimentación final para el alumno.

   Consejo 5. Revisa el video para conocer más sobre el uso del portafolio
   https://www.youtube.com/watch?v=-_ov41oNPu8

Formato de acompañamiento para la evaluación

1) **Seleccione el instrumento que acompañará la evaluación del portafolio.**
   Puede utilizar una rúbrica, lista de cotejo u otros instrumentos.
2) **Retome los criterios de evaluación.**
   Puede considerar como criterios la organización y calidad de las evidencias, el uso efectivo del medio y el desarrollo del portafolio.

   Consejo 6. Se recomienda que el instrumento que utilice incluya un apartado de observaciones para que pueda registrar elementos fundamentales del proceso que desarrolló el alumno al integrar el portafolio.

3) **Diseñe su instrumento de evaluación.**
   Para diseñarlo puede consultar los capítulos sobre los instrumentos: rúbrica y lista de cotejo.

Revisión

- Una vez que utilice el portafolio, es muy útil que pida a los alumnos comentarios o sugerencias tanto del portafolio, como de la evaluación que realizó del mismo.
- Un ejercicio para enriquecer el portafolio y el instrumento complementario, es compartirlos con un grupo de expertos (otros profesores) para recibir sugerencias, percepciones, debates y consensos que surjan de la retroalimentación.

¿CÓMO LO APLICO?

- El portafolio es un instrumento que, mediante las evidencias o testimonios que compendia, permite visualizar el proceso de aprendizaje del alumno y se recomienda que su desarrollo sea durante todo el curso de formación.
- Es sustancial que defina y comunique cuáles son las características que espera observar en las evidencias para que sus alumnos las conozcan y logren cumplir con sus expectativas.

  Consejo 7. El análisis inicial de los objetivos, contenidos y actividades lo guiará en este proceso.

- Determine la periodicidad con la que revisará las evidencias del portafolio, es importante que establezca sesiones de seguimiento del desarrollo del portafolio con la finalidad de generar un espacio de retroalimentación para el alumno donde identifique los problemas y necesidades.
- Las evidencias que los alumnos pueden incluir varían, según la naturaleza del espacio educativo. Además, debe asegurarse que representen un desafío y un elemento de motivación para los alumnos.

  Consejo 8. Las evidencias pueden ser individuales, pero si la planeación de su clase lo requiere pueden ser en equipo.

EL PORTAFOLIO EN LA EVALUACIÓN A DISTANCIA

¿Cómo lo diseño?

Para el diseño del portafolio en la evaluación a distancia la planeación mantiene un proceso similar, sin embargo, deberá considerar ciertas implicaciones derivadas de realizarlo a distancia, entre las cuales se encuentran:

- Examinar el medio tecnológico en el que el alumno presentará el portafolio.
- Verificar que el medio que seleccionó es compatible con su espacio de trabajo virtual y que domina su uso para atender cualquier problema que surja.
- Contemplar el tiempo que le llevará al alumno conformar su portafolio en el medio seleccionado, por ejemplo, generar audios, elaborar videos o integrar presentaciones y a su vez, almacenarlos en una plataforma o medio seleccionado.
- Contemplar que los alumnos pueden tener acceso limitado a infraestructura, *software* y *hardware*. Bajo este escenario, considere actividades que el alumno pueda desarrollar en su contexto para darle una oportunidad de aprendizaje y evitar su rezago.

  Consejo 9. En algunas ocasiones el uso de plataformas o medios populares que utilizan los alumnos frecuentemente en otras clases o redes sociales pueden ser una opción.

- Fomentar retroalimentaciones sincrónicas y asincrónicas con sus alumnos. Las primeras sirven para atender dudas emergentes y promueven la presencia social y la socialización del aprendizaje, mientras que las segundas proporcionan una guía que el estudiante puede utilizar en el momento en que emprende su proceso de aprendizaje de manera independiente.

¿Cómo lo aplico?

- El portafolio puede emplearse en distintas plataformas o aplicaciones, pero es importante que la selección se realice con base en:
  - El tiempo del que disponen los estudiantes para crear su portafolio.
  - La disposición de recursos para implementarla, es decir, si las herramientas requieren de conexión a internet o *software* especializado, opte por aquéllas que son de acceso libre.
  - El tipo de supervisión durante la creación del portafolio, por ejemplo, si será el proyecto final del curso deberá decidir si tendrá reuniones periódicas para ver el avance de la creación del portafolio.
- Logre comunicar adecuadamente los pasos que seguirá el alumno para la creación de su portafolio en la herramienta que seleccione.
- Previo a la implementación, es indispensable que realice una prueba de la herramienta que utilizará desde el rol del estudiante, para verificar la vista del usuario, la redacción y gramática del contenido, el tiempo del que se dispone y su correcto funcionamiento.
- Recuerde que el uso de cualquier herramienta requiere de una guía sobre su funcionamiento para los alumnos como usuarios, para que su desempeño no se vea afectado. Por ello, es importante que usted la conozca desde el rol del alumno.
- Existe una gran cantidad de herramientas que puede adaptar para el portafolio a distancia, recuerde que algunas son gratuitas y esto permite potenciar su creatividad.

¿CÓMO ANALIZO SUS RESULTADOS?

- El análisis de los resultados del portafolio permite valorar el nivel de dominio de los contenidos y permite apreciar el crecimiento y desarrollo personal del alumno o del grupo de alumnos conforme a su trayectoria de aprendizaje.
- Para utilizar el portafolio se recomienda el uso de un instrumento que acompañe la evaluación como una rúbrica o lista de cotejo, pues permite definir cuáles son los aspectos que desea valorar y el nivel de desempeño que espera en cada uno.
- El instrumento de acompañamiento le permitirá identificar en las evidencias los aspectos a evaluar y el nivel de desempeño que presentan. Con el análisis de la información de las evidencias podrá visualizar el nivel de desempeño de cada uno de sus alumnos y las problemáticas del proceso de enseñanza, con ello, podrá evaluar al estudiante y mejorar su práctica docente.

- El uso de otro instrumento de acompañamiento para la evaluación del portafolio permite al profesor y al alumno, conocer los contenidos y criterios que serán analizados para determinar su aprendizaje. De esta forma, ambos pueden orientar sus estrategias educativas para cumplir con los aspectos y niveles deseables de la formación.
- A nivel individual, usted puede visualizar el proceso de aprendizaje que experimentó cada uno de los alumnos.
- A nivel de equipos de trabajo, puede conocer los desafíos, logros y estilos, que, en conjunto, presentaron los alumnos y de esta forma, identificar las áreas que requieren su atención para fortalecer su proceso de enseñanza.
- A nivel generacional, usted puede comparar el desempeño de las diferentes generaciones que han desarrollado el portafolio y generar preguntas que le permitan comparar estas experiencias para mejorar su práctica docente.

EJEMPLOS

Ejemplo 1

Formato para portafolio estructurado en físico



Asignatura:

Objetivo del portafolio: Valorar la producción de trabajos académicos realizados durante el semestre

Formato del portafolio: Físico

Tipo del portafolio: Estructurado

Fecha de entrega: Mayo

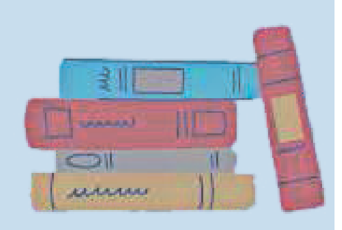
Instrucciones

Integra un portafolio con las siguientes secciones.

En cada una debes incluir las evidencias estipuladas. Deberás presentarlo en una carpeta con el diseño visual que elijas.

Secciones

Sección 1
Sección 2
Sección 3



Evidencias

De la sección 1 se deben incluir las evidencias:

Evidencia A
Evidencia B
Evidencia C
Evidencia D

De la sección 2 se deben incluir las evidencias:
Evidencia E
Evidencia F

De la sección 3 se deben incluir las evidencias:
Evidencia G

Calendario de retroalimentaciones

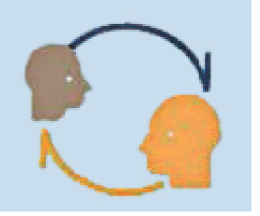
Sesión de retroalimentación 1 -enero- 2 horas

Sesión de retroalimentación 2 -febrero- 2 horas

Sesión de retroalimentación 3 -marzo- 2 horas

Retroalimentación final

Junio - 4 horas



Formato de acompañamiento para la evaluación del portafolio estructurado en físico

Rúbrica

Nombre del evaluado: ___________________________________ Fecha: _______________

A continuación, se presentan cinco criterios para la evaluación de un portafolio estructurado en físico, en cada uno se describen cuatro niveles de ejecución con un puntaje. En la última columna indique el puntaje que considera que el portafolio alcanza.

Criterios	Sobresaliente 20	Suficiente 17.5	Deficiente 12.5	Inaceptable 10	Puntaje
Secciones	Presenta todas las secciones requeridas.	Presenta más de la mitad de las secciones requeridas.	Presenta menos de la mitad de las secciones requeridas.	No presenta las secciones requeridas.	
Evidencias	Incluye todos los tipos de evidencias solicitados que respaldan el logro del aprendizaje esperado: tareas y actividades en cada sección.	Incluye más de la mitad de los tipos de evidencias solicitados que respaldan el logro del aprendizaje esperado: tareas y actividades en cada sección.	Incluye menos de la mitad de los tipos de evidencias solicitadas y no demuestran el logro del aprendizaje esperado: tareas y actividades en algunas secciones.	No incluye los tipos de evidencias solicitados ni demuestran el logro del aprendizaje esperado: tareas y actividades en algunas secciones.	
Organización de evidencias	Las evidencias cumplen con la secuencia establecida y usa un formato creativo.	Las evidencias tienen una secuencia coherente que no es la establecida, pero usa un formato creativo.	Las evidencias tienen una secuencia poco clara, no es la establecida y usa un formato poco creativo.	Las evidencias no tienen una secuencia coherente y usa un formato poco creativo.	

Criterios	Sobresaliente 20	Suficiente 17.5	Deficiente 12.5	Inaceptable 10	Puntaje
Estado de las evidencias	Las evidencias demuestran los avances en los aprendizajes esperados y cumplen con todos los elementos del formato que se consideran en las instrucciones del portafolio.	Las evidencias demuestran los avances en los aprendizajes esperados, pero no cumplen con todos los elementos del formato que se consideran en las instrucciones del portafolio.	Las evidencias demuestran algunos de los avances en los aprendizajes esperados y cumplen con algunos de los elementos del formato que se consideran en las instrucciones del portafolio.	Las evidencias no demuestran los avances en los aprendizajes esperados, pero cumplen con algunos de los elementos del formato que se consideran en las instrucciones del portafolio.	
Reflexiones	Contiene reflexiones serias y vinculadas con los logros alcanzados y los aspectos para mejorar en cada sección.	Contiene reflexiones serias y poco vinculadas con los logros alcanzados y los aspectos para mejorar en cada sección.	Contiene reflexiones poco vinculadas con los logros alcanzados y los aspectos para mejorar en algunas secciones.	No contiene reflexiones sobre los logros alcanzados ni los aspectos para mejorar en ninguna de las secciones.	
				Total	

Nombre del evaluador: ____

Fuente: Elaboración propia

Ejemplo 2

Formato para portafolio semiestructurado en línea

Asignatura:
Objetivo del portafolio: Valorar la producción de trabajos académicos realizados durante el semestre
Formato del portafolio: En línea **Tipo de portafolio:** Semiestructurado

Instrucciones	Secciones
Integra un portafolio con secciones que consideres adecuadas. En cada una debes incluir las evidencias obligatorias y en el caso de las optativas, elige las que consideres representativas de tu aprendizaje. Deberás presentarlo utilizando la herramienta Digication (www.digication.com).	El alumno tiene la libertad de establecer las secciones y el orden de sus evidencias

Evidencias

Obligatorias Es indispensable que incluyas		Optativas Secciona una de cada tipo	
Evidencia A	☐	**Evidencia d**	☐
Evidencia B	☐	**Evidencia f**	☐
Evidencia C	☐	**Evidencia g**	☐

Calendario de retroalimentación

Sesión de retroalimentación 1 - enero - 2 horas
Sesión de retroalimentación 2 - febrero - 2 horas
Sesión de retroalimentación 3 - marzo - 2 horas

Fecla de entrega	Retroalimentación final
Mayo	Junio - 4 horas

Formato de acompañamiento para la evaluación del portafolio semiestructurado en línea

Registro de observaciones

Nombre del evaluado: ______________________ Fecha: ____________________________

El siguiente registro de observaciones contiene seis elementos guía, en los cuales puede describir las fortalezas y/o dificultades que percibió en el portafolio semiestructurado en digital.

Organización y clasificación de las secciones

____
____
____
____

Presentación de evidencias obligatorias

____
____
____
____

Selección de evidencias optativas

____
____
____
____

Estado de las evidencias obligatorias y optativas

____
____
____
____

Uso de la herramienta digital

____
____
____
____

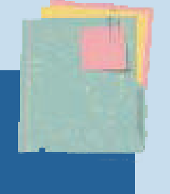
Reflexiones vinculadas con los logros alcanzados y los aspectos para mejorar en cada sección

____
____
____
____

Nombre del evaluador: ____

Fuente: Elaboración propia

Ejemplo 3
Formato para portafolio libre en línea



Asignatura:
Objetivo del portafolio: Valorar la producción realizada en el semestre
Tipo de portafolio: Libre
Formato del portafolio: En línea

Instrucciones:

Selecciona los trabajos que a tu juicio consideres los mejores productos que has realizado durante el semestre. En cada uno incluye una reflexión sobre cada trabajo. Deberás presentarlo utilizando la herramienta Mahara (www.mahara.org).



Secciones:

Tienes la libertad de establecer las secciones y el orden de tus trabajos



Calendario de retroalimentación:

Sesión 1 - enero - 2 horas

Sesión 2 - febrero - 2 horas

Sesión 3 - marzo - 2 horas

Final - junio - 4 horas



Evidencias:

Tienes la libertad de seleccionar los trabajos que mostrarás como evidencia de tu aprendizaje



Fecha de entrega: Mayo





Formato de acompañamiento para la evaluación del portafolio libre en línea

Escala de valoración

Nombre del alumno: ________________________ Fecha: __________________________
A continuación, se presentan seis criterios para la evaluación del portafolio. Indique el puntaje obtenido en la propuesta del alumno.

Evaluación para el portafolio de evidencias	Puntaje	Puntaje del alumno
El alumno demuestra dominio de las actividades realizadas.	15	
El alumno consideró las indicaciones en las actividades realizadas.	15	
El alumno demuestra conocimiento de los procesos realizados.	20	
Hay evidencia de su progreso en el desarrollo de las actividades.	10	
La autoevaluación de sus trabajos demuestra un proceso de reflexión constante y serio.	30	
El alumno asistió a las sesiones de retroalimentación.	10	
	Total	
Observaciones		

Nombre del evaluador: ____

Fuente: Elaboración propia

Formato de autoevaluación

Nombre del alumno: ______________________ Fecha: ________________________

Instrucciones. Contesta las siguientes preguntas reflexionando sobre cada uno de los trabajos que analizaste, seleccionaste y organizaste para construir tu portafolio.

1.- ¿Por qué seleccionaste este trabajo como uno de los mejores? ¿Qué criterio utilizaste?

____

____

2.- Al realizar el trabajo, ¿qué fue lo más importante para ti?

____

____

3.- ¿En qué es similar o diferente a los otros trabajos que seleccionaste?

____

____

4.- ¿Cuáles son, según tu criterio, tus áreas fuertes? ¿Cuáles son las débiles?

____

____

5.- Si pudieras volver a realizar el portafolio ¿Cambiarías algo?

____

____

Fuente: Elaboración propia.

RECURSOS EN LÍNEA

- Es importante considerar que, gracias a las numerosas aportaciones del portafolio a los procesos de enseñanza, aprendizaje y evaluación profesionales en diversas disciplinas han logrado adaptarlo de manera fácil y rápida a medios digitales, es decir, que una nueva modalidad es integrarlo de manera digital pero manteniendo su esencia en cuanto a que siga representando una selección analítica de trabajos desarrollados tanto por los estudiantes, con productos derivados de su formación, como por los docentes creando un espacio digital para compartir sus reflexiones y experiencias frente al acto educativo. En este sentido, los términos portafolio digital, portafolio electrónico y E-portafolio son distintas expresiones para hacer referencia al portafolio dispuesto en un medio digital (Cortés, Pinto y Atrio, 2015).
- El E-Portafolio refiere a la implementación de un portafolio en un entorno digital constituido por evidencias y actividades virtuales como físicas que son presentadas por medio de algunas Tecnologías de la Información y la Comunicación (TIC). Esta recopilación digital promueve la construcción del conocimiento y fomenta el desarrollo de competencias digitales (Cortés, Pinto y Atrio, 2015). Derivado de este concepto actualmente se pueden aprovechar recursos en línea de los cuales se han creado nuevos conceptos entre los cuales se encuentran:

- **M-Portafolio:** es un portafolio digital desarrollado exclusivamente por medio de aplicaciones móviles.
- **Netfolio:** es una red de un conjunto de portafolios electrónicos producidos por los alumnos que, por medio de la comunicación en línea, brinda nuevo contenido y diferentes perspectivas bajo un enfoque de colaboración posibilitando la comparación de los portafolios como una etapa más de la construcción individual del portafolio personal (Barberá-Gregori, 2008).
- **Transfolio:** está diseñado para realizar procesos de evaluación de competencias transversales, donde el alumno va construyendo y compartiendo su identidad académica, así como la profesional. Las evidencias se valoran y se pasa por un proceso de negociación y diálogo con el profesor para evaluar los distintos niveles de adquisición de competencias.
- **B-Portafolio:** es un portafolio digital, basado exclusivamente en plataformas de *blogging* y redes sociales.

Derivados de los conceptos anteriores, se han desarrollado recursos digitales que aumentan la diversidad de evidencias que se pueden integrar en un portafolio, algunas de ellas son:

Recurso	Descripción
Mahara http://mahara.org	Es un sistema independiente que permite tener un ambiente virtual de aprendizaje centrado en el alumno, además es una aplicación hermana de Moodle y cuenta con una aplicación móvil. Permite recibir comentarios públicos y privados, se pueden crear grupos y cuenta con un marco de evidencia que permite visualizar el progreso de cada usuario.
Foliotek http://folioteck.com/	Es un ***software*** con diferentes perfiles de compra y una prueba gratis de siete días. Tiene una variedad de herramientas que le permiten al usuario realizar evaluaciones, crear portafolios y mostrar proyectos a empleadores.
Eduportfolio http://eduportfolio.org	Es una plataforma que facilita la organización, el intercambio y la búsqueda de información. Está disponible en siete idiomas y permite presentar la información en forma de vitrinas, cuenta con seis herramientas que facilitan administrarlas.
Digication https://www.digication.com/	Es un ***software*** fácil y simple de usar que permite combinar herramientas de intercambio de archivos y publicación en línea. Además, los usuarios pueden crear una comunidad para aprender dentro y fuera del aula. Cuenta con diferentes planes de precios que incluyen actualizaciones, respaldo, almacenamiento ilimitado, sesiones de capacitación para su uso y acceso a estándares y rúbricas de evaluación.

CONCLUSIONES Y REFLEXIONES FINALES

- El uso del portafolio ayuda al profesor y al alumno a planificar y gestionar los procesos de enseñanza-aprendizaje.

  Consejo 10. Te invitamos a ver el siguiente video donde encontrarás una recopilación de la información que te hemos presentado hasta el momento: https://www.coursera.org/lecture/evaluacion-cualitativa/portafolio-y-tipos-de-portafolio-84kM2

- Permite visualizar el aprendizaje de los alumnos durante el ciclo de formación.
- Apoya a los alumnos a autoevaluar su propio trabajo y a identificar sus logros y áreas de oportunidad.
- El portafolio es un instrumento que demanda una planeación minuciosa para facilitar la evaluación con los instrumentos de acompañamiento.
- Cuando los grupos son numerosos se recomienda la implementación de pequeños equipos de trabajo y la coevaluación entre estos. Así, el profesor puede supervisar el trabajo de los grupos de una forma más práctica.

REFERENCIAS

Barberá-Gregori, E. (2005). Calificar el aprendizaje mediante la evaluación por portafolios. *Perspectiva Educacional, Formación de Profesores, 45*, 70-84. Recuperado de http:// www.redalyc.org/articulo.oa?id=333329100005

Barberá-Gregori, E. (2008). Mutual feedback in e-portfolio assessment: an approach to the netfolio system. *British Journal of Educational Technology*, *40*, 342-357. Recuperado de https://onlinelibrary.wiley.com/doi/full/10.1111/j.1467-8535.2007.00803.x

Barberá-Gregori, E. (2009). Portafolios electrónicos y educación superior en España: Situación y tendencias. *Revista de Educación a Distancia,* (8), 1-13. Recuperado de https://www.um.es/ead/red/M8/intro.pdf

Barragán, R., García, R., Buzón, O., Rebollo, M. y Vega, L. (2009). E-Portafolio en Procesos Blended-Learning: Innovaciones de la Evaluación en los Créditos Europeos, *Revista de Educación a Distancia, 8*, 216.

Cortés, O., Pinto, A., Atrio, S. (2015). E-portafolio como herramienta construccionista del aprendizaje activo en tecnología educativa. *Revista Lasallista de Investigación*. *12*(2), 36-44. Recuperado de http://www.scielo.org.co/scielo.php?pid=S1794-44492015000200004&script=sci_abstract&tlng=es

Díaz-Barriga, F., Martínez-Romero, E. y Heredia-Sánchez, A. (2012). Diseño tecnopedagógico de portafolios electrónicos de aprendizaje: Una experiencia con estudiantes universitarios. *Revista Electrónica de Investigación Educativa, 14*(2), 103-107. Recuperado de http://www.redalyc.org/articulo.oa?id=15525013008

Font-Ribas, A. (2004). Líneas maestras del aprendizaje por problemas, *Interuniversitaria de Formación del Profesorado, 18*(1), 79-95. Recuperado de http://www.redalyc.org/articulo.oa?id=27418106

García-Martínez, B. y Baena-Escudero, R. (2009). El uso del portafolios en la docencia universitaria: experiencia de renovación metodológica en la asignatura de Geomorfología Fluvial (titulación de Geografía) *Revista Iberoamericana de Educación, 3*(49), 1-11. Recuperado de http://rieoei.org/2703.htm

Gutiérrez-Morales, G. y Sánchez-Mendiola, M. (2020). El uso del portafolio electrónico en evaluación educativa en Medicina. *Revista Española de Educación Médica*. (2), 65-73. Recuperado de https://revistas.um.es/edumed/article/view/445671

Martínez-González, A. y Soto-Hernánadez, G. (2021). *Evaluación para el aprendizaje: enfoque cualitativo. Portafolio* [MOOC] https://www.coursera.org/lecture/evaluacion-cualitativa/portafolio-y-tipos-de-portafolio-84kM2

Martínez-Sánchez, N. (2002). El portafolio como mecanismo de validación de aprendizaje. *Perfiles educativos, 24*(95), 54-66. Recuperado de http://www.redalyc.org/articulo.oa?id=13209505

Capítulo 22

DEMOSTRACIÓN

María Juliana Londoño Cárdenas, María Elena Pérez Rivera

"You can't use up creativity. The more you use, the more you have."
Maya Angelou

INTRODUCCIÓN

La demostración es una técnica empleada tanto para enseñar como para evaluar habilidades, herramientas y aprendizajes específicos. De acuerdo con sus características no solo permite la comprensión profunda acerca de un tema, sino que suscita el interés en los estudiantes (DeKorver, Choi y Towns, 2017).

La evaluación, mediante la demostración, requiere estar alineada con el marco curricular y la enseñanza para que ésta tenga sentido y significado tanto para el docente como para el estudiante. Dicha técnica pertenece a los métodos cualitativos de evaluación y se inscribe dentro de la evaluación auténtica que busca valorar el desempeño de los estudiantes a partir de situaciones reales y problemas significativos complejos (Ahumada-Acevedo, 2005).

La evaluación de una demostración a distancia presenta varios retos, sin embargo, es la más significativa para comprobar que los estudiantes adquirieron el conocimiento y observar el modelo mental en el que se incorpora información visual, espacial, temporal y conceptual (Lowe, Boucheix y Fillisch, 2017). Además, al presentarse en línea, los recursos disponibles para la elaboración de demostraciones son más accesibles e innovadores que al realizarlos de forma presencial.

El presente capítulo se centrará en el uso de la demostración como técnica de evaluación del y para el aprendizaje. Su objetivo es presentar de forma clara y sencilla la información necesaria para emplearla. Está constituido por ocho secciones en forma de preguntas, la primera define qué es una demostración, la segunda detalla cada uno de los pasos para diseñarla, la tercera recomienda cómo aplicarla, la cuarta considera elementos para ser utilizada en la modalidad a distancia, la quinta sugiere cómo analizar los resultados, la sexta presenta tres ejemplos, la séptima contiene recursos en línea y la octava incluye algunas conclusiones y reflexiones finales.

¿QUÉ ES?

- Es una técnica que permite evaluar habilidades y conocimientos específicos del estudiante, la cual debe llevarse a cabo en espacios educativos públicos. Implica que el estudiante exponga, explique o aplique, ante el profesor y una audiencia particular, el procedimiento, el proceso de un tema o el tópico bajo estudio, en forma concreta. Es decir, mediante una demostración un estudiante realiza una ejecución real o simulada ante otros.
- Con la demostración se puede valorar la apropiación, comprensión o capacidad de aplicación de una teoría, método, técnica o algún instrumento por parte del estudiante. Además de poder apreciar la definición propia de conceptos, actitudes y habilidades relacionadas con la resolución de problemas, el pensamiento crítico, la comunicación efectiva. Esto permite involucrar al alumno como monitor de su propio aprendizaje, todo ello fomenta la metacognición (Deese, Ramsey, Walczyk y Eddy, 2000).
- Para elaborarla es necesaria una colaboración multidisciplinaria, ya que implica habilidades y conocimientos de distintas asignaturas. Por ejemplo, se ha empleado en formaciones que demandan dominio e interrelación, como en los campos de la química, física, matemática, ingeniería, informática, música, teatro y danza, entre otros. Lo que permite que se aplique en diferentes niveles educativos.
- En esta técnica se pueden considerar demostraciones teóricas o prácticas, en las cuales los estudiantes trabajan juntos para recopilar sus hallazgos, con el fin de poder discutir qué han aprendido de acuerdo con la información que se les proporcionó durante la demostración.
- Esta técnica se puede presentar durante o al final del curso. Las actividades solicitadas deberán incluir trabajos que ayuden al estudiante a aplicar lo aprendido, predecir fenómenos relacionados o expandir su conocimiento general de lo estudiado.
- Una forma de demostración, es la interactiva la cual permite impartir un tema específico y evaluar formativamente la comprensión de los estudiantes. El siguiente ejemplo se basa en su secuencia:

Figura 1. Ejemplo de una demostración interactiva

Basado en Harizah, Kusairi y Latifah. (2020).

- La evaluación por medio de la demostración está diseñada a partir de los resultados que deben lograr los estudiantes. En esta evaluación, el estudiante es un participante activo.
- Ésta es utilizada para evaluar el logro de los aprendizajes esperados, en los cuales se convierten abstracciones complejas en aplicaciones reales o reproducciones de un ejemplo, logrando así reconocer las fortalezas y debilidades del estudiante.

¿CÓMO LA DISEÑO?

Para realizar una evaluación, basada en demostración, se necesita hacer un análisis del mapa curricular, del perfil de egreso y del programa de asignatura, para decidir si se evaluarán solamente uno o varios temas utilizando la técnica de demostración.

Planeación

1) **Determine los aprendizajes esperados que evaluará.**
   Con base en el análisis realizado anteriormente, ¿cuáles son los aprendizajes que pretendo valorar con la demostración?

Consejo 1. Al elaborar los aprendizajes esperados considere incluir los conceptos claves para la búsqueda de la información y el desarrollo de la actividad 1 de este apartado.

2) **Precise el tipo de evaluación que realizará.**
¿Esta evaluación tiene un fin diagnóstico, formativo o sumativo?
3) **Establezca quiénes participarán en la evaluación.**
¿Se realizará una autoevaluación, heteroevaluación o coevaluación? Considere que puede emplear una o todas las anteriores.
4) **Defina las características de la demostración.**
¿Se realizará en un escenario real, simulado o virtual? ¿Cuánto tiempo dispone para el desarrollo de una demostración? Considere un tiempo aproximado en el que los participantes demuestren soluciones a un problema tomando en cuenta la naturaleza, la profundidad y la complejidad del tema.
5) **Escoja si será por equipo o individual.**
¿El desarrollo de la demostración se hará a nivel individual o grupal? En esta técnica es necesario que, si hay una demostración en equipo, se evalúe a cada uno de los participantes.

Consejo 2. Tome en cuenta que la demostración implica cierta complejidad y tiempo, por lo que se recomienda un trabajo colaborativo.

6) **Seleccione el producto que utilizará.**
Escoja uno o varios productos que permitan valorar el objetivo que se planteó y que estén acorde con el tema. Éstos pueden ser documentos, presentaciones o exposiciones, una elaboración de un procedimiento mediante una maqueta, entre otros.
7) **Defina con qué criterios evaluará la demostración.**
¿Qué aspectos evaluará con la demostración? ¿Cuáles serán los indicadores observables? ¿Qué peso le otorgará a cada indicador? Es importante tomar en cuenta no solo el contenido sino también cómo se ejecuta.
8) **Elija el instrumento que utilizará para la evaluación.**
¿Qué instrumento facilita la recopilación de información acerca del nivel de desempeño de los estudiantes y el logro de los aprendizajes? Por ejemplo, listas de cotejo, rúbricas y registro de observación, entre otros.

Se sugiere la siguiente actividad para llevar a cabo la planeación. Complete la Tabla 1, tomando en cuenta la relación lógica presente entre los conceptos clave y las actividades que desee proponer como técnica de evaluación.

Tabla 1. Actividad

Conceptos clave de los aprendizajes esperados	Actividades de demostración	Tiempo destinado a las actividades	Escenario real / simulado / virtual	Individual o grupal	Productos para evaluar	Instrumentos de evaluación

Diseño

1) **Presente el problema o el enunciado.**
Exponga de manera sencilla y clara tanto el contenido, problema o enunciado que los estudiantes deberán desarrollar y los tiempos de entrega por actividad o etapa, como los instrumentos que el profesor va a utilizar para recabar información y los criterios e indicadores de evaluación.

Consejo 3. Entregue esta presentación por escrito. Puede elaborar una lista de cotejo con los materiales que se requieren para elaborar esta actividad.

2) **Proponga llevar una bitácora de trabajo.**
Presente una propuesta para llevar una bitácora de trabajo en donde se especifiquen los pasos, los cambios y las dudas que se puedan presentar durante el proceso de la demostración.
3) **Asesore la preparación y el desarrollo de la demostración.**
Sugiera bibliografía relevante sobre el tema específico, además de programar sesiones con los estudiantes para dar retroalimentaciones oportunas para que éstos puedan mejorar su demostración. Es necesario que ensaye la demostración previamente, con el fin de detectar algún paso en el que los estudiantes se puedan confundir y así poder asesorarlos de una mejor manera.
4) **Seleccione los instrumentos de evaluación.**
Con base en los aprendizajes esperados, elabore un instrumento que le permita evidenciarlos. Éstos pueden ser rúbrica, lista de cotejo, entre otros. Para elaborarlos puede ayudarse de los manuales de cada instrumento.
5) **Comunique los resultados de la evaluación.**
Es importante que la calificación se entregue junto con una retroalimentación del trabajo, para que ésta ayude al estudiante a detectar qué habilidades debe mejorar y en cuáles excede expectativas.

6) **Considere una retroalimentación final.**
Planee un espacio en el que se haga una retroalimentación general del proceso, en el que se expresen los progresos y las áreas en las que deben trabajar cada uno de los estudiantes.

Revisión

- Es necesario analizar periódicamente los avances de cada una de las etapas de la demostración para retroalimentar a los estudiantes.
- Al ser de carácter interdisciplinario, se debe tener en cuenta que es necesario integrar todas las evidencias de trabajo del estudiante y revisarlas individualmente; en caso contrario es posible que se tenga una visión fragmentada de lo que se pretende evaluar.
- Se debe considerar evaluar la creatividad y la individualidad de cada trabajo, ya que son elementos fundamentales al momento de elaborarlo. En el caso de las demostraciones elaboradas en equipo, se sugiere recopilar evidencias a nivel individual, por ejemplo, en un portafolio.

¿CÓMO LA APLICO?

- Antes de evaluar con una demostración, es necesario que, dentro de las actividades de aprendizaje, se incluyan demostraciones prácticas y teóricas, de tal manera que el estudiante considere tanto los recursos que se requieren, como el procedimiento a seguir para una demostración, es decir, orden, claridad y precisión.
- La demostración sirve para evaluar el aprendizaje del estudiante durante el proceso enseñanza-aprendizaje, así como al término de éste, ya que muestra evidencia sobre el dominio de un contenido específico.
- Esta técnica puede abarcar una amplia variedad de proyectos, presentaciones, actividades o productos. A continuación, se mencionan algunos:

    - Presentaciones orales.
    - Documentales o *podcast.*
    - Poemas y cuentos.
    - Experimentos científicos, uso de analogías para explicar fenómenos, estudios y reportes.
    - Elaboración de modelos, esculturas, dioramas, robots, etcétera.
    - Una presentación.
    - Ejecuciones en escenarios reales o simulados.

- La evaluación debe orientarse hacia la retroalimentación, ya que, al ser oportuna puede conducir a los estudiantes a ajustar el planteamiento de la demostración.
- Este tipo de evaluación requiere, por parte del docente y del estudiante, una fuerte inversión de tiempo, por lo que es necesario considerar los espacios disponibles para la revisión de las evidencias, la retroalimentación y la carga de trabajo de los participantes en otras materias (dedicación).

- Se sugiere que las observaciones del proyecto, presentación, actividad o producto se realicen de manera pública, para que puedan ser consideradas por cada uno de los estudiantes, y con base en ello no cometer los mismos errores. Se deberá realizar en un espacio de respeto, reflexión y orientación hacia el otro.

LA DEMOSTRACIÓN EN LA EVALUACIÓN A DISTANCIA

Al realizar una evaluación a distancia, basada en demostración, es necesario seguir los pasos presentados en los apartados: "¿Cómo la diseño?" y "¿Cómo lo aplico?", además de las siguientes recomendaciones.

¿Cómo la diseño a distancia?

- Elabore un banco de recursos en línea en donde recopile sitios, aplicaciones, blogs, entre otros, que puedan ser de ayuda para el desarrollo de la demostración y que sean de acceso abierto para todos los participantes. Puede pedir sugerencias a sus compañeros, y valorar la pertinencia y accesibilidad.
- Al solicitar una demostración en línea, deberá brindar instrucciones que contengan indicaciones puntuales del contenido, el recurso audiovisual o audio a emplear, y las especificaciones del recurso como: peso, calidad, formato, duración, entre otros.
- Al determinar los materiales que requiere para la elaboración del producto, tome en cuenta la disponibilidad y accesibilidad que tengan los estudiantes.
- Como actividad de evaluación diagnostica incluya una demostración realizada por usted y luego proponga que los estudiantes compartan su explicación del fenómeno presentado. Considere que los materiales que utilice sean visualmente atractivos y que el procedimiento, contenido o concepto sea claro. Por ejemplo: puede usar envases etiquetados, pizarrones en línea o materiales de colores contrastantes.
- Al explicar la demostración en vivo asegure la grabación de ésta ya que le servirá para sesiones asíncronas futuras o material de consulta. En caso que no cuente con los recursos para elaborar una demostración en vivo, apóyese en videos o tutoriales que sirvan para explicar el contenido.
- Al realizar sesiones asincrónicas podrá apoyarse de recursos como: *Edpuzzle*[1], *Vialogues*, *TEDEd*[2], entre otros. Se sugiere elaborar blogs o cuestionarios para recopilar opiniones y/o respuestas de los estudiantes sobre los recursos propuestos.

---

1 Edpuzzle ES. (15 de marzo de 2021). Google Classroom y Edpuzzle [Archivo de Vídeo]. Youtube. https://www.youtube.com/watch?v=QpN1yuZ2-60

2 Ayudantes UC Temuco. (20 de diciembre de 2016). ¿Cómo usar TEDEd? [Archivo de Vídeo]. Youtube. https://www.youtube.com/watch?v=6Pg_yrlzY5M
Unidad de Universidad Abierta UDA Azuay. (7 de junio de 2020). Uso de herramientas: TEDEd [Archivo de Vídeo]. Youtube. https://www.youtube.com/watch?v=pvB06flJrOs

- Si requiere que los estudiantes presenten una demostración en vivo se recomienda grabar la sesión para que quede evidencia del trabajo y pueda ser revisada con detalle en un momento posterior por los estudiantes y el profesor.

¿Cómo lo instrumento?

- Para incrementar el compromiso de los estudiantes se sugiere incluir actividades de discusión y coevaluación orientadas hacia la mejora del trabajo y comprensión del tema que pueden estar alojadas en un blog, LMS, Google Drive, entre otros.
- Programe sesiones regulares con los estudiantes que le permita conocer el nivel de comprensión del contenido visto y definir si requieren una actividad o recursos adicionales para potenciar el aprendizaje. Estas sesiones se pueden llevar a cabo de manera sincrónica o asincrónica en redes sociales (por medio de grupos privados), correo electrónico, LMS, chats, entre otros.
- Conozca y aprópiese de los materiales o recursos que proponga a los estudiantes, esto impactará en el interés y uso que hagan de ellos. Además, asegure que el material sea atractivo para todos y pertinente para los objetivos de aprendizaje esperados.
- Entregue retroalimentación efectiva cuidando que el contenido sea respetuoso y significativo para el aprendizaje del estudiante. Si emplea retroalimentación oral, puede utilizar notas de audio, videos cortos o sesiones sincrónicas; y si se utiliza la retroalimentación escrita use los foros del LMS, correo electrónico o la función de comentarios en Word, PDF, PowerPoint, Excel, entre otros.
- En las sesiones sincrónicas proponga actividades de evaluación que puedan realizarse de manera rápida y les den pistas a los estudiantes para saber si se está comprendiendo el fenómeno o procedimiento demostrado.
- Registre las sesiones sincrónicas para tener evidencias acerca del nivel de aprendizaje de cada estudiante.
- Utilice distintos instrumentos para evaluar el proceso de la demostración y no solo el producto final, por ejemplo, durante el desarrollo proporcionar a los estudiantes una rúbrica que describa el desempeño esperado y oriente la evaluación por pares.

¿CÓMO LO ANALIZO?

- Con cada uno de los criterios establecidos para evaluar la demostración, se deberá elaborar una retroalimentación constructiva que ayude al estudiante a conocer sus fortalezas y las áreas de oportunidad en las que debe mejorar su desempeño.
- A partir del cumplimiento de los criterios e indicadores se pueden planear actividades de apoyo a nivel grupal e individual.
- A continuación, se presentan algunos tipos de evaluación que pueden servir de apoyo para valorar el desempeño de los estudiantes:

- ❍ **Portafolio:** recopilar evidencias sobre todo el proceso implicado en la demostración, por ejemplo, la bitácora, los protocolos, la lista de materiales, entre otros.
- ❍ **Cuestionario semiestructurado:** seleccionar preguntas que permitan evidenciar la articulación de ideas y conocimientos aprendidos.
- ❍ **Rúbrica:** elegir los criterios y el peso que se le dará a cada uno.
- ❍ **Evaluación por pares:** elaborar una lista de cotejo con los criterios de evaluación y solicitar la valoración por parte de los compañeros.
- ❍ **Autoevaluación:** definir una serie de preguntas para el estudiante que propicien la reflexión del tema.

- Las herramientas mencionadas anteriormente mostrarán las particularidades que presentaron los estudiantes durante el desarrollo de la demostración. La información recopilada servirá para poder orientar al estudiante acerca de los contenidos que le resultaron más complicados de entender y demostrar y, a su vez, para detectar qué temas domina.

EJEMPLOS

1. Rúbrica para evaluar una demostración teórica

Instrucción: Marque el nivel alcanzado por el estudiante en cada criterio considerando la descripción propuesta. En caso de que una sección o un criterio no se aplique al tipo de reporte por favor señale con una marca la columna "No aplica".

	Bueno	Regular	Necesita revisión	No aplica
Tesis	Responde al tema. Contiene un argumento principal comprensible y claro.	Responde al tema parcialmente. Contiene un argumento principal comprensible, pero no claro.	No responde al tema. El argumento principal no es comprensible ni claro.	
Argumentos de apoyo	Los argumentos de apoyo y citas presentados soportan el argumento principal. Emplea la bibliografía correctamente para sustentar los argumentos.	Los argumentos de apoyo y citas presentados soportan parcialmente el argumento principal. Emplea la bibliografía parcialmente para sustentar los argumentos.	Los argumentos de apoyo y citas presentados no soportan el argumento principal. Emplea la bibliografía, pero no sustenta los argumentos.	

	Bueno	Regular	Necesita revisión	No aplica
Análisis	Los conceptos se entienden con precisión. Las ideas presentadas tienen un orden lógico.	Los conceptos son parcialmente claros. Las ideas presentadas tienen un orden, pero éste no es lógico.	Los conceptos no son claros. Las ideas presentadas no cuentan con un orden ni son lógicas.	
Conclusiones	Las conclusiones presentadas están relacionadas con lo expuesto en el documento. Son coherentes y comprensibles. Dan un cierre al tema presentado.	Las conclusiones presentadas se relacionan parcialmente con lo expuesto en el documento. Son comprensibles, pero no coherentes. Dan un cierre incompleto al tema presentado.	Las conclusiones presentadas no se relacionan con lo expuesto en el documento. No son coherentes ni comprensibles. No dan cierre al tema presentado.	
Ortografía	No contiene faltas de ortografía. El lenguaje utilizado es claro y apropiado para la disciplina.	Contiene menos de cinco faltas de ortografía. El lenguaje utilizado es vago y no apropiado para la disciplina.	Contiene faltas de ortografía. No contiene evidencias de un uso de lenguaje apropiado para la disciplina.	
Fuentes	Se cita, menciona y referencia correctamente.	Se referencia correctamente, pero no se cita y menciona correctamente.	No cita, menciona o referencia correctamente.	

Fuente: Elaboración propia

2. Autoevaluación de un estudiante sobre su reporte de una demostración práctica mediante una lista de cotejo

Instrucciones. A partir del trabajo que has realizado, marca con una "X" si consideras que cuenta o no con los elementos mencionados.

Estudiante: ______________________________

Tema de la demostración: ______________________________

El reporte de mi demostración cuenta con:

Título	Sí	No
Introducción	Sí	No
Fuentes de información	Sí	No
Objetivos	Sí	No
Marco teórico	Sí	No

Planteamiento del problema	Sí	No
Desarrollo del problema	Sí	No
Relación del problema con el producto	Sí	No
Diseño del producto	Sí	No
Especificaciones del producto	Sí	No
Lista del material que utilicé	Sí	No
Evidencias del desarrollo del producto	Sí	No
Conclusiones	Sí	No
Bibliografía	Sí	No

Observaciones

____

____

____

____

____

Fuente: Elaboración propia

3. Coevaluación de una demostración práctica

Instrucciones: De acuerdo con la demostración de tu compañero, contesta el siguiente cuestionario de forma respetuosa y honesta; considera que las siguientes preguntas le ayudarán a mejorar su aprendizaje.

Tabla 3.

Nombre del expositor:	**Fecha:**	
Tema de la demostración:	**Duración:**	
Definió correctamente el tema	SÍ	NO
¿Por qué?		
Explicó correctamente las ideas principales	SÍ	NO
¿Por qué?		
Los ejemplos demostrados fueron pertinentes	SÍ	NO
¿Por qué?		
Los ejemplos fueron claros	SÍ	NO

¿Por qué?		
Los ejemplos fueron interesantes	SÍ	NO
¿Por qué?		
Me permitió aclarar el tema	SÍ	NO
¿Por qué?		
La forma en que fue planteada me resultó innovadora	SÍ	NO
¿Por qué?		
La manera en que se desarrolló me resultó creativa	SÍ	NO
¿Por qué?		
Observaciones		
Recomendaciones		

RECURSOS EN LÍNEA

Recurso	Descripción
Rúbrica para demostraciones https://gc.scalahed.com/recursos/files/r156r/w13868w/R%C3%BAbrica_demostraciones.pdf	Es un archivo pdf que contiene una rúbrica que puede ser empleada para evaluar una demostración práctica.
The Coalition of Essential Schools http://www.essentialschools.org/resources/	Ésta es una organización que tiene un rol central en el desarrollo de demostraciones prácticas en escuelas. El sitio contiene diversos artículos y videos en los que se observan el desarrollo y la evaluación de una demostración.
New York Performance Standards Consortium https://www.performanceassessment.org	Es una página orientada a presentar ejemplos de rúbricas en distintas áreas de conocimiento, además incluye una muestra de trabajos de estudiantes.
Proyecto Aprender http://proyectoaprender.es/formacion/es-ES/recursos/aplicaciones-educativas	Sitio en el que se muestran aplicaciones educativas que pueden servir como herramienta para el proceso de evaluación de una demostración.
The Learning Accelerator https://practices.learningaccelerator.org/problem-of-practice/remote-friendly-assessments-how-can-i-begin-building-assessments-that-are-accessible-for-students-in-remote-and-hybrid-learning-environments	Artículo que explica cómo se pueden construir evaluaciones accesibles para los estudiantes en ambientes de aprendizaje remotos e híbridos.

CONCLUSIONES Y REFLEXIONES FINALES

- La demostración reconoce los aprendizajes que no pueden ser valorados por una evaluación tradicional en distintos campos de conocimiento, ya que implica que el estudiante relacione sus conocimientos previos con una comprensión profunda de un tema en específico para apropiarse del tema y poder expresarlo claramente de forma verbal o escrita.
- Fortalece la idea del trabajo interdisciplinario en el sentido que los estudiantes utilizan herramientas que son propias de otras áreas para resolver problemas que surjan de su demostración.
- A nivel individual y grupal, favorece el desarrollo de habilidades como argumentación, asertividad, creatividad y construcción de conceptos; permitiendo al profesor tener una idea del modelo mental que está adquiriendo el o los estudiantes sobre un tema en específico lo que facilitará la selección de recursos didácticos que faciliten la apropiación de los contenidos.
- Es importante utilizar distintos instrumentos como la bitácora de trabajo, lista de cotejo, diario de campo, entre otros, como insumo para la autoevaluación.
- Las demostraciones cortas sucesivas afianzan los aprendizajes esperados, porque permiten el logro y dominio de los conocimientos de los estudiantes.

REFERENCIAS

Ahumada-Acevedo, P. (2005) La Evaluación auténtica: un sistema para la obtención de evidencias y vivencias de los aprendizajes. *Perspectiva Educacional, formación de profesores*, 11-24. https://www.redalyc.org/articulo.oa?id=333329100002

Abbott, S., Guisbond, L., Levy, J. y Sommerfeld, M. (2014). *The glossary of education reform*. https://www.edglossary.org/

Basheer, A., Hugerat, M., Kortam, N. y Hofstein, A. (2017) The Effectiveness of Teachers' Use of Demonstrations for Enhancing Students' Understanding of and Attitudes to Learning the Oxidation-Reduction Concept. *EURASIA Journal of Mathematics Science and Technology Education*. https://doi.org/10.12973/eurasia.2017.00632a

Bowen, C. W. y Phelps, A. J. (1997). Demonstration-based cooperative testing in general chemistry: A broader assessment-of-learning technique. *Journal of Chemical Education*, *74*(6), 715–719. https://doi.org/10.1021/ed074p715

Davidson, J. (2009) Exhibitions: Connecting Classroom Assessment with Culminating Demonstrations of Mastery, *Theory into practice*, *48*(1), 36-43.

DeKorver, B., Choi, M. y Towns, M. (2017). Exploration of a Method to Assess Children's Understandings of a Phenomenon after Viewing a Demonstration Show. *Journal of Chemical Education*, *94*, 149−156. https://doi.org/10.1021/acs.jchemed.6b00506

Deese, W., Ramsey, L., Walczyk, J. y Eddy, D. (2000). Using Demonstration Assessments to Improve Learning. *Journal of Chemical Education*, 77 (11), 1511-1516. https://doi.org/10.1021/ed077p1511

Harizah, Z., Kusairi, S., y Latifah, E. (2020) Student's critical thinking skills in interactive demonstration learning with web based formative assessment. *Journal of Physics: Conference Series*, *1567*(4). https://doi.org/10.1088/1742-6596/1567/4/042038

Lowe, R., Boucheix, J-M. y Fillisch, B. (2017) Demonstration tasks for assessment. En: Lowe, R. y Ploetzner, R. (Ed.) *Learning from Dynamic Visualization. Innovations in Research and Application*. 177-201. https://doi.org/10.1007/978-3-319-56204-9_8

Matos, Y. y Pasek, E. (2008) La observación, discusión y demostración: técnicas de investigación en el Aula. *Laurus, 14*(27), 33-52. http://www.redalyc.org/articulo.oa?id=76111892003

Ontario Ministry of Education (2002). *The Ontario Curriculum Unit Planner: Assessment Strategies Companion*. Teacher Companion. Toronto: The Ministry.

Pierce, D. y Pierce, T. (2007). Effective use of Demonstration Assessment in the Classroom Relative to Laboratory Topics. *Journal of Chemical Education, 84*(7). https://doi.org/10.1021/ed084p1150

Pittí, K., Curto-Diego, B. y Moreno-Rodilla, V. (2010). Experiencias construccionistas con robótica educativa en el centro internacional de tecnologías avanzadas. *Teoría de la Educación. Educación y Cultura en la Sociedad de la Información, 11*(1), 310-329. http://www.redalyc.org/articulo.oa?id=201014897013

Capítulo 23

EXPOSICIÓN ORAL

María Azucena Montoya Magno, Nancy Fabiola Pérez Herrera, Ana Laura Pérez Díaz

"El que sabe pensar, pero no sabe expresar lo que piensa,
está en el mismo nivel que el que no sabe pensar."
PERICLES

INTRODUCCIÓN

La universidad como espacio de formación académica debe contribuir en el desarrollo de competencias que les sean útiles a los estudiantes, no solo a lo largo de su formación, sino también aquellas que le sirvan para afrontar cualquier desafío en el futuro (Verano et al., 2016).

Algunos autores mencionan que el nivel medio superior y superior son espacios para promover aprendizajes disciplinares y competencias transversales básicas tales como la lectura, la escritura y la expresión oral (Teruel et al., 2020). La comunidad escolar tiene que centrarse en facilitar su desarrollo en diferentes espacios educativos, más que en probar su importancia para el éxito profesional (Verano et al, 2016; Cooper et al., 2021).

Así, la *comunicación oral* es una de las competencias más valoradas en el ejercicio de cualquier actividad profesional, por lo que, pese a la basta cantidad de definiciones existentes y a la dificultad para encontrar la mejor, sin duda es una de las competencias vitales para las actividades profesionales futuras, en una amplia variedad de espacios laborales (Arce-Medina, 2006; Costigan y Brink, 2020; Verano et al., 2016).

La exposición, también llamada presentación oral, disertación o conferencia, es una técnica didáctica muy utilizada en los espacios educativos y esto tiene que ver con que gran parte del proceso formativo requiere del diálogo académico, el debate y la argumentación, por lo que es un medio para el desarrollo de la competencia de comunicación oral (Verano et al., 2016), en la cual también se ve reflejado el razonamiento crítico y la capacidad de análisis y de síntesis, consideradas como habilidades para la vida.

La exposición oral puede ser empleada por el profesor como una técnica didáctica de enseñanza y como un recurso de evaluación del aprendizaje de los alumnos. En ambos casos, la *evaluación de la exposición oral* debe considerar aspectos que van desde el tipo de

argumento que se utiliza hasta los recursos materiales empleados como apoyo (Cooper et al., 2021; Costigan y Brink, 2020; Verano et al., 2016).

Cuando el profesor utiliza la exposición oral, como recurso de evaluación del y para el aprendizaje de sus alumnos, debe planear, organizar y establecer los criterios a considerar en la evaluación de manera precisa, con el propósito de que esta técnica coadyuve a mejorar el proceso de enseñanza-aprendizaje. En este sentido, la evaluación juega un papel preponderante al concebirla como todo aquello que contribuya a que los procesos de construcción del conocimiento de los estudiantes mejoren y que permite que se retroalimente al alumno de manera constante (Pimienta-Prieto, 2008), además de que sus resultados se usan para la toma de decisiones sobre los planes futuros de enseñanza o de aprendizaje (Yan et al., 2021).

De ahí, la necesidad de presentar a los profesores diversas herramientas que le permitan evaluar las exposiciones orales con el objetivo de potenciar su uso, ya sea como técnica didáctica de enseñanza o como recurso de evaluación del y para el aprendizaje.

En este capítulo se aborda la exposición oral como una de las técnicas que pueden apoyar la evaluación del y para el aprendizaje. Su propósito es proporcionar a los profesores una herramienta que les ayude en el diseño y uso de la misma, está organizado en seis secciones: en la primera se explica qué es, en la segunda se detallan los pasos para diseñarla, en la tercera se recomienda cómo aplicarla, además de que en esta sección se presenta información del desarrollo de la técnica en un marco de evaluación a distancia, para ello se especifica cómo diseñarla y aplicarla en dicha modalidad. En la cuarta se puntualiza cómo analizar sus resultados (ya sea en modalidad presencial, a distancia o híbrida). La quinta sección presenta ejemplos y recursos en línea y, por último, en la sexta se presentan conclusiones y reflexiones finales.

¿QUÉ ES?

- La exposición oral es una técnica que consiste en la presentación de un tema a una audiencia.
- En el proceso formativo la exposición oral requiere del diálogo académico, el debate y la argumentación.
- Puede llevarse a cabo en dos formas: individual o grupal y en diferentes modalidades: presencial, a distancia o híbrida.
- La exposición oral es una técnica muy utilizada en el nivel medio superior y superior. A pesar de que su uso no es exclusivo de un área educativa en particular, se sabe que en las Ciencias Sociales, Humanidades y Artes es muy usada.
- Es útil en la evaluación formativa porque se puede valorar si el alumno es capaz de buscar, organizar, analizar y sintetizar información, establecer relaciones entre contenidos y comunicarlos a una audiencia de manera fluida y coherente. El seguimiento del proceso de preparación de la exposición brinda la oportunidad tanto al profesor como a los compañeros de retroalimentar el trabajo del alumno o alumnos que van a exponer un tema.
- Mediante la exposición se pueden evaluar conocimientos tales como la identificación de términos, conceptos o hechos relevantes y habilidades de análisis y síntesis de

información, dimensiones de la comunicación oral, así como trabajo en equipo, manejo de grupos, entre otras.

- Una de las ventajas del uso de la exposición oral como técnica para la evaluación del y para el aprendizaje es que permite evaluar una amplia gama de contenidos en un tiempo relativamente corto, esto depende del propósito para la que fue diseñada y por el cual, el profesor se decidió por el uso de dicha técnica.
- Una de las limitaciones en el uso de la exposición oral como recurso de evaluación del y para el aprendizaje es enfrentarse a grupos muy numerosos y no lograr los objetivos para la que fue diseñada; si se abusa en el uso de esta técnica puede resultar cansada y monótona para los alumnos.

Figura 1. ¿Qué es la exposición oral?



¿CÓMO LA DISEÑO?

La exposición oral, utilizada como una técnica de evaluación del y para el aprendizaje, requiere de una serie de consideraciones en su planeación para garantizar que cumplirá con el propósito de la evaluación. Es importante asegurar que la enseñanza, el aprendizaje y la evaluación estén alineados, por lo que es fundamental analizar el mapa curricular en el que se encuentra inserta la asignatura, así como el perfil de egreso y el programa de la asignatura, esto le ayudará a identificar los aprendizajes esperados que son susceptibles a evaluarse por medio de la exposición.

Planeación

1) **Seleccione los aprendizajes esperados que se evaluarán por medio de la exposición.**
   Se recomienda que reflexione sobre las siguientes preguntas: ¿Estos aprendizajes esperados se pueden evaluar por medio de una exposición?, ¿Es viable dadas las características del grupo y el tiempo del que se dispone?, ¿Cuál es la experiencia que tienen los alumnos en el uso de esta técnica?
2) **Determine el tipo de evaluación.**
   La exposición es útil en la evaluación formativa ya que le permite retroalimentar el proceso de preparación de la exposición y que los alumnos realicen los ajustes necesarios para que se alcancen los aprendizajes esperados. En el caso de la evaluación sumativa, le permitirá identificar el logro de los aprendizajes esperados al término de un tema, una unidad o un curso.
3) **Especifique quiénes serán los participantes en la evaluación.**
   Tradicionalmente, el profesor ha sido el actor principal en la evaluación de la exposición oral (heteroevaluación); sin embargo, en la actualidad se hace énfasis en la participación activa de los alumnos en su aprendizaje, lo que también implica que estén involucrados en su propia evaluación (autoevaluación) y en la de sus pares (coevaluación). En resumen, la exposición es una técnica que permite la participación del profesor y de los alumnos en la evaluación.
4) **Defina las características de la exposición.**
   Forma (individual o grupal), modalidad (presencial, a distancia o híbrida) etapas (planeación, acompañamiento y ejecución) y recursos utilizados en formato presencial: carteles, mapas mentales, presentaciones PowerPoint, Genially, Prezi, HaikuDeck o Canva, modelos o videos, en formato a distancia: plataforma de comunicación síncrona o asíncrona (Zoom, Cisco Webex Meetings, Skype, Google Hangouts, Google Meet, Facetime, GoToMeeting, Flipgrid, entre otros).
   Organice junto con los alumnos la logística de la exposición; determinen el calendario de presentación y los recursos que serán necesarios para garantizar que se lleven a cabo en tiempo y forma.
5) **Determine los criterios de evaluación con base en los aprendizajes esperados.**
   Elabore criterios de evaluación claros y precisos en donde especifique qué espera que demuestren los alumnos de forma individual: Dominio del tema, habilidades de análisis y síntesis, habilidades de comunicación oral, manejo de grupo, uso de materiales de apoyo, fuentes de información consultadas, relación del discurso con las imágenes, apoyo en el material escrito, tonalidad y modulación de la voz, claridad al hablar, expresión corporal, elocuencia, gestión o control del tiempo, entre otros.

Cuando se trata de una actividad grupal o en equipos: aclarar los aspectos que tienen que ver con el trabajo colaborativo previo al momento de la exposición, asimismo, durante la exposición se puede valorar la coordinación entre los integrantes, la calidad de las diapositivas utilizadas, el orden, secuencia, coherencia y claridad de la presentación, entre otros.

Figura 2



6) **Elija los instrumentos que acompañarán la evaluación de la exposición.**
Para que la evaluación de la exposición se lleve a cabo de una manera sistemática, se recomienda el uso de instrumentos tales como la rúbrica, la lista de cotejo y la escala de calificación que apoyan el proceso de evaluación (para mayores detalles sobre estos instrumentos consulte los capítulos 14. Lista de cotejo o 15. Rúbrica, de este libro).

Diseño

1) Prepare el formato del esquema general que deberá tener la exposición y entregue o envíe a los alumnos.
2) Pregunte si hay dudas o comentarios acerca de las características que deberá presentar la exposición.
3) Entregue el calendario de exposiciones.
4) Presente a los alumnos los instrumentos que se utilizarán para evaluar el desempeño de la exposición.
5) Informe quiénes serán los participantes en la evaluación: heteroevaluación (profesor), coevaluación (compañeros) y/o autoevaluación.
6) Prepare los materiales que necesitará en cada una de las exposiciones (lista de asistencia, formato de retroalimentación y de evaluación).
7) Tome algunas fotografías o realice una grabación durante el desarrollo de la exposición que sirvan como evidencia de la actividad (con previa autorización de los alumnos).

Revisión

- Es recomendable que se realicen revisiones periódicas del avance en la preparación de la exposición, el profesor debe asegurarse que los alumnos cumplan con lo solicitado en el perfil de la exposición. Es decir, en este tipo de evaluación no es dejar que los alumnos preparen solos la exposición oral, es necesario acompañarlos durante la preparación de la misma.

- Proporcione retroalimentación a los alumnos respecto al trabajo realizado con el propósito de hacer los ajustes necesarios antes de la exposición. Si el desempeño no fue el deseado ofrezca opciones de entrenamiento para hablar frente al público como: grabar algún podcast, hablar frente al espejo, grabarse mediante plataformas digitales de entrenamiento para empoderar su voz y fortalecer la confianza para comunicar ideas, entre otras.

¿CÓMO LA APLICO?

- La exposición oral puede llevarse a cabo en distintos escenarios, por ejemplo, en el espacio educativo, un laboratorio, un auditorio o un espacio abierto; sin embargo, es necesario que las condiciones y los recursos requeridos sean considerados desde la planeación.
- La modalidad de la exposición oral puede ser individual o grupal. Es importante que desde la planeación considere qué aprendizaje quiere evaluar, para así orientar su decisión respecto al número de integrantes que conformará el equipo.
- Es importante respetar los tiempos asignados a cada expositor o equipo de expositores, así como el número de exposiciones que se programaron en las fechas establecidas.
- La exposición oral puede programarse para evaluar un tema al inicio, durante o al final de un curso.
- Se recomienda tener listos los formatos con los criterios de evaluación del desarrollo de la exposición oral.
- Al finalizar la exposición o exposiciones orales programadas considere un espacio para brindar retroalimentación al o a los expositores. En la retroalimentación se sugiere mencionar los aspectos que el alumno domina y aquellos en los que debe mejorar.

La exposición oral en evaluación a distancia

El contexto actual, ha impulsado a los sistemas educativos y a los docentes a transitar hacia modelos o esquemas de trabajo a distancia que requieren ir más allá de los procesos tradicionales de enseñanza-aprendizaje presenciales. Este cambio implica soluciones mediadas por la tecnología que pueden ofrecer oportunidades de práctica, independientemente del tiempo y lugar, que favorecen el desarrollo de habilidades asociadas a la competencia oral (Boetje y Ginkel, 2020).

Así, por ejemplo, en los últimos meses se han utilizado diversas aplicaciones informáticas mediante las cuales es posible realizar videoconferencias grupales síncronas: Zoom, Cisco Webex Meetings, Skype, Google Hangouts, Google Meet, Facetime, GoToMeeting, entre otras. Si bien para el desarrollo de la exposición oral es usual recurrir a alguna de estas aplicaciones, en caso de que no sea posible llevar a cabo la reunión virtual en un mismo momento, la exposición oral también se puede desarrollar a través de medios asíncronos como las videograbaciones.

Lo más importante de resaltar es que se ha identificado que la práctica de las presentaciones orales, con o sin apoyo de tecnología, tiene grandes beneficios para el desarrollo de la

competencia oral, ya que ayuda a reducir la ansiedad o aprehensión comunicativa y, en consecuencia, mejorar la ejecución de la presentación oral (Boetje y Ginkel, 2020).

¿Cómo la diseño a distancia?

Independientemente de que la exposición oral se utilice en modalidad presencial, a distancia o híbrida, su diseño requiere de un proceso de planificación cuidadoso, con sentido formativo y orientado a propósitos específicos y a alcanzar la alineación o correspondencia entre estos componentes al momento de planificarlos.

Si la exposición oral se lleva a cabo mediante una videoconferencia síncrona, se sugiere idear secuencias de actividades que impliquen diversos momentos para las presentaciones orales, ya que se ha observado un mayor beneficio a las secuencias con presentaciones orales únicas; es decir, la calidad de la exposición oral aumenta si se utiliza como una estrategia continua de trabajo en clase. Al contrario, si no es posible el desarrollo de una videoconferencia síncrona, también son de utilidad los diversos tipos de entrenamiento para la exposición, por ejemplo: videograbaciones individuales, ensayo frente a un espejo, etc.

Además, dentro de la exposición considere una sesión de preguntas y respuestas para demostrar la capacidad de los alumnos para formular argumentos, pero también de escucha, ya que esta es una habilidad intelectual de gran importancia en el desarrollo de la comunicación oral (Costigan y Brink, 2020).

Diseño de la rúbrica de calificación para la exposición oral en educación a distancia

Si la exposición oral es vista como técnica de evaluación ya sea individual o grupal, y se utiliza una rúbrica para evaluarla, diséñela para orientar a los alumnos en el nivel de logro que se desea alcanzar, es decir, incluya los criterios y descripciones que serán evaluados y compártalos previamente. Recuerde que lo importante es definir los atributos y que estos sean congruentes con el aprendizaje esperado y que sean observables durante la exposición (individual o grupal). Para conocer con mayor detalle el proceso de elaboración de la rúbrica consulte el capítulo 15 de este libro.

Consulte el video "Exposición oral." Youtube. https://youtu.be/oNSp4xDLVcA

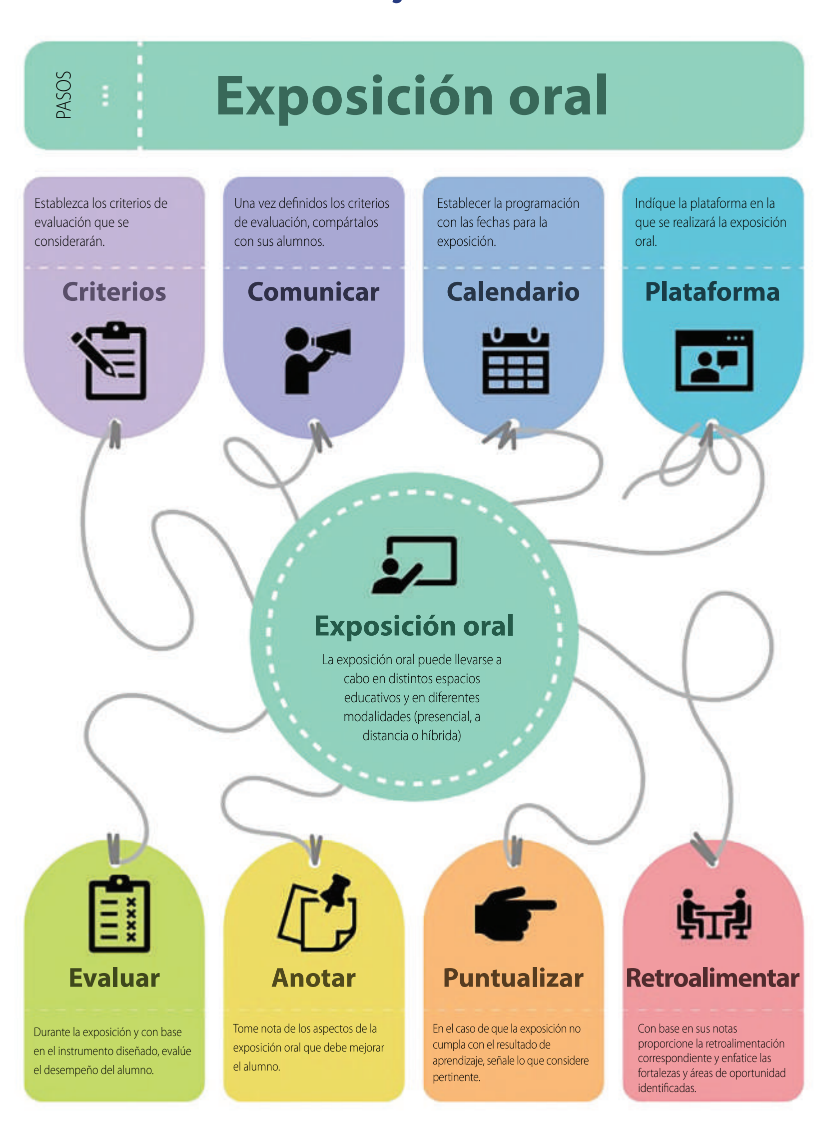
¿Cómo la instrumento?

- La exposición oral puede llevarse a cabo en distintos espacios educativos y en diferentes modalidades (presencial, a distancia o híbrida).
- Si se trabaja en modalidad a distancia considere el uso de alguna plataforma para llevar a cabo la exposición.
- Comunique a sus alumnos la fecha y hora en que se realizará la exposición.
- Defina los criterios de evaluación que utilizará para valorar el desempeño de los alumnos durante la exposición.
- Se recomienda tener listos los formatos con los criterios de evaluación del desarrollo de la exposición oral. Pueden ser plantillas en computadora como, por ejemplo, un archivo

de Excel con las descripciones de los criterios y los puntajes asignados para facilitar la suma de puntos de cada equipo o a través de alguna herramienta tecnológica (páginas de rúbricas en línea). En cualquier caso, integre un apartado final para incluir observaciones generales de la exposición oral que le serán útiles al momento de compartir la retroalimentación a los alumnos.

- Programe una sesión para proporcionar la retroalimentación a sus alumnos en donde indique las fortalezas y áreas de oportunidad identificadas.
- Anime a sus alumnos a realizar ejercicios para ensayar sus habilidades de comunicación.

Figura 3



¿CÓMO ANALIZO SUS RESULTADOS?

- Para evaluar la exposición oral, ya sea en modalidad presencial, a distancia o híbrida, es necesario establecer previamente los criterios que se considerarán en el desempeño de los alumnos; dichos criterios dependerán del propósito de la evaluación. Por lo anterior, se puede inferir que los criterios de evaluación son diversos y, por lo tanto, no existe un modelo único para llevarlo a cabo; sin embargo, es fundamental asegurarse que la evaluación de una exposición oral se lleve a cabo de manera sistemática.
- Los criterios que el profesor seleccione para evaluar una exposición oral, deben estar definidos claramente desde la planeación y ser del conocimiento de los alumnos, con el objetivo de que, tanto el profesor como el alumno, manejen la misma información y entiendan el proceso de evaluación.
- Dado que la exposición oral implica que los alumnos pongan en marcha diferentes habilidades y que es posible evaluarlas en etapas distintas del desarrollo de la exposición (planeación, acompañamiento y ejecución), en la Tabla 1, a manera de guía se presenta una propuesta de indicadores de evaluación de una exposición oral agrupados por categorías (tema a desarrollar, habilidades o recursos de apoyo) y por etapas en las que se puede valorar el desempeño de los alumnos. Por ejemplo, si considera que un aspecto a valorar en el desempeño del alumno tiene que ver con la secuencia, coherencia y claridad del contenido de la exposición, este criterio es factible valorarlo previo a la ejecución, esto es, desde la planificación que el alumno realiza de la exposición, pero también durante el acompañamiento que proporciona al alumno en el desarrollo de la actividad y por supuesto, en el momento de la ejecución. La elección de los indicadores dependerá de los aspectos que desea evaluar en el alumno en relación con la exposición oral.

Tabla 1. Indicadores para evaluar la exposición oral agrupados por categoría y etapa de evaluación

Indicadores	Etapa		
	Planificación	Acompañamiento	Ejecución
TEMA A DESARROLLAR			
Contenido de la exposición (secuencia, coherencia y claridad)	*	*	*
Organización del contenido de la exposición (secuencia, coherencia y claridad)	*	*	*
Dominio del tema			*
Calidad de los argumentos expresados			*
Bibliografía y fuentes consultadas (confiables y actuales)	*	*	

HABILIDADES			
Capacidad de análisis y síntesis de información	*	*	
Hablar frente a la audiencia (tonalidad y modulación de voz, claridad, fluidez y elocuencia)			*
Expresión corporal (postura, gesticulaciones, movimiento de manos, contacto visual con la audiencia)			*
Dominio del espacio			*
Gestión y control de tiempo	*	*	*
Manejo de grupo			*
Autorregulación (planificación, monitoreo, reacción y reflexión)	*	*	*
Toma de decisiones	*	*	*
Capacidad de escucha y formulación de argumentos ante preguntas específicas			*
RECURSOS DE APOYO			
Relación del discurso con los medios o recursos de apoyo			*
Manejo de los medios o recursos de apoyo			*
Calidad visual de los medios o recursos de apoyo		*	*
SI SE TRATA DE UNA ACTIVIDAD GRUPAL			
Coordinación entre los integrantes del grupo (trabajo colaborativo, comunicación, orden y secuencia en la presentación)	*	*	*

Fuente: Elaboración propia

Nota: La tabla es ilustrativa, el profesor puede determinar aquellos indicadores para evaluar la exposición oral que sean más congruentes con el objetivo del uso de esta técnica.
Según la modalidad (presencial, a distancia o híbrida) podrán hacerse ajustes a los indicadores de evaluación de la exposición oral.

- Una vez que se tengan definidos los criterios, se recomienda seleccionar el instrumento que acompañará la evaluación de la exposición oral. Algunos instrumentos sugeridos para llevar dicha evaluación son: la rúbrica, la lista de cotejo o la escala de calificación.
- El análisis de los resultados dependerá del instrumento que el profesor seleccione; sin embargo, se sugiere que, además de establecer un puntaje de calificación (Excelente=4, Satisfactorio=3, Suficiente=2, Deficiente=1), se acompañe de un análisis cualitativo del desempeño del alumno durante la exposición (interacción con la audiencia, autenticidad y entusiasmo), lo anterior independientemente de la modalidad de trabajo (presencial, a distancia o híbrida).

EJEMPLOS

Tabla 2. Lista de cotejo para la evaluación de una exposición oral individual o grupal

Nombre del expositor o expositores:	
Tema:	
Objetivo de aprendizaje:	

Instrucciones: Indique la presencia o ausencia de los siguientes criterios en el desempeño de la exposición oral del alumno o alumnos.

Sí	No	Criterios de evaluación del desempeño en una exposición oral individual
☐	☐	El contenido de la exposición es coherente y claro.
☐	☐	Domina el tema de exposición.
☐	☐	Utiliza argumentos de calidad.
☐	☐	Analiza y sintetiza la información.
☐	☐	Habla de manera clara y fluida frente a la audiencia.
☐	☐	La expresión corporal es adecuada (postura, gesticulaciones, movimientos de mano y contacto visual).
☐	☐	Gestiona y controla el tiempo de exposición.
☐	☐	Propicia un ambiente adecuado (capta la atención de sus compañeros).
☐	☐	Existe relación del discurso con los recursos de apoyo que utiliza.
☐	☐	Maneja adecuadamente los medios o recursos de apoyo.
☐	☐	La calidad visual de los medios o recursos de apoyo es buena.
Sí	**No**	**Criterios de evaluación del desempeño en una exposición oral grupal**
☐	☐	El contenido de la exposición es coherente y claro.
☐	☐	Dominan el tema de exposición.
☐	☐	Utilizan argumentos de calidad.
☐	☐	Analizan y sintetizan la información.
☐	☐	Hablan de manera clara y fluida frente a la audiencia.
☐	☐	La expresión corporal es adecuada (postura, gesticulaciones, movimientos de mano y contacto visual).

☐	☐	Gestionan y controlan el tiempo de exposición.
☐	☐	Propician un ambiente adecuado (captan la atención de sus compañeros).
☐	☐	Se observa trabajo en equipo durante la exposición.
☐	☐	Existe coordinación entre los integrantes del equipo al momento de hablar.
☐	☐	Respetan una secuencia durante la exposición.
☐	☐	Existe relación del discurso con los recursos de apoyo que utilizan.
☐	☐	Manejan adecuadamente los medios o recursos de apoyo.
☐	☐	La calidad visual de los medios o recursos de apoyo es buena.
Observaciones		

Fuente: Elaboración propia.

Tabla 3. Rúbrica para evaluar una exposición

Instrucciones: A continuación, se presenta un ejemplo con algunos aspectos a evaluar en la exposición oral. Asigne el puntaje correspondiente al desempeño del alumno cuando se trate de una actividad individual o de los alumnos si se trata de una actividad grupal.

Aspectos a evaluar en la exposición oral	Excelente (4 puntos)	Satisfactorio (3 puntos)	Suficiente (2 puntos)	Deficiente (1 punto)
TEMA A DESARROLLAR				
Organización del contenido de la información	El contenido de la exposición muestra una secuencia lógica y coherente. Las ideas presentadas son claras y precisas.	El contenido de la exposición muestra una secuencia lógica y coherente. Sin embargo, algunas de las ideas presentadas son ambiguas e imprecisas.	El contenido de la exposición en algunos casos no muestra una secuencia lógica y coherente. Algunas de las ideas presentadas son ambiguas e imprecisas.	El contenido de la exposición no muestra una secuencia lógica y coherente. La mayoría de las ideas presentadas son ambiguas.
Bibliografía y fuentes consultadas	Los textos o artículos consultados corresponden a fuentes de información confiables y actuales. Tienen relación con el tema que se aborda y corresponden a publicaciones de cinco años anteriores a la fecha.	Los textos o artículos consultados corresponden a fuentes de información confiables. Tienen relación con el tema que se aborda y corresponden a publicaciones de más de cinco años anteriores a la fecha.	Algunos de los textos o artículos consultados corresponden a fuentes de información poco confiables. Tienen relación con el tema que se aborda y corresponden a publicaciones de más de cinco años anteriores a la fecha.	La mayoría de los textos o artículos consultados corresponden a fuentes de información poco confiables. Tienen relación con el tema que se aborda, pero corresponden a publicaciones de más de cinco años anteriores a la fecha.
HABILIDADES				
Capacidad de análisis y síntesis de la información	El expositor presenta la información de manera ordenada y lógica (enfatiza cuáles son las ideas principales y las complementarias, utiliza conocimientos previos, resume la información presentada, construye un punto de vista a partir de los comentarios generados durante la exposición).	El expositor presenta la información de manera ordenada (enfatiza cuáles son las ideas principales, utiliza conocimientos previos y resume la información presentada durante la exposición).	El expositor presenta la información de manera ordenada (enfatiza cuáles son las ideas principales y resume la información presentada durante la exposición).	El expositor no presenta la información de manera ordenada (no enfatiza cuáles son las ideas principales y no es capaz de resumir la información presentada durante la exposición).

Aspectos a evaluar en la exposición oral	Excelente (4 puntos)	Satisfactorio (3 puntos)	Suficiente (2 puntos)	Deficiente (1 punto)
Manejo de grupo	El expositor es quien lleva la sesión, modera la participación de sus compañeros. La intervención del profesor es mínima.	El expositor es quien lleva la sesión, modera medianamente la participación de sus compañeros. La intervención del profesor es media.	El expositor es quien lleva la sesión, pero no modera la participación de sus compañeros. La intervención del profesor es alta.	El expositor no es capaz de llevar la sesión, no modera la participación de sus compañeros, el profesor tiene que intervenir en el orden durante toda la exposición. La intervención del profesor es muy alta.
RECURSOS DE APOYO				
Relación del discurso con los medios o recursos de apoyo	Los medios o recursos de apoyo se emplean para ejemplificar o reforzar los argumentos expresados en la exposición.	Los medios o recursos de apoyo son utilizados únicamente para aclarar conceptos.	Algunos de los medios o recursos de apoyo contribuyen a la exposición, pero no son tan efectivos.	Se emplean pocos medios o recursos de apoyo y no son acertados para el contenido de la exposición.
Calidad visual de los medios o recursos de apoyo	Los medios o recursos de apoyo son visualmente atractivos, interactivos y contienen ideas fundamentales del tema.	Los medios o recursos de apoyo reflejan organización, aunque visualmente no son atractivos ni interactivos y contienen ideas generales del tema.	Los medios o recursos de apoyo presentan sobrecarga de información y deficiencias en su construcción y su relación con el tema es difusa.	Los medios o recursos de apoyo presentan una estructura que dificulta la lectura, no son interactivos y se alejan por completo del tema de la exposición.
SI SE TRATA DE UNA ACTIVIDAD GRUPAL				
Coordinación entre los integrantes del grupo	La participación de los integrantes del grupo refleja una planificación. Todos participan en la exposición y los cambios de expositor se realizan con una transición lineal del tema.	La participación de la mayoría de los integrantes refleja una planificación. Todos participan en la exposición, pero en algunos cambios de expositor se observa variación en la información, lo que impide mantener la secuencia del tema.	Solo la participación de algunos integrantes muestra el seguimiento de una planificación. Todos participan en la exposición, pero en la mayoría de los cambios de expositor no se sigue la secuencia del tema.	La presentación de los integrantes está desarticulada y se observa falta de colaboración. Solo algunos miembros del grupo participan en la exposición.
OBSERVACIONES GENERALES DEL DESEMPEÑO EN LA EXPOSICIÓN				
Este espacio está destinado para integrar comentarios generales que le serán de utilidad al momento de proporcionar retroalimentación a los alumnos.				

Fuente: Elaboración propia.

RECURSOS EN LÍNEA

(Herramientas de evaluación a distancia que puedan apoyar al docente para generar y aplicar el instrumento)

En la actualidad existe un gran número de herramientas tecnológicas que mejoran los recursos de apoyo que se utilizan para la exposición oral. A continuación, se presenta un breve listado para ilustrar la amplia gama de posibilidades existentes y que puede sugerir a sus estudiantes con el propósito de elaborar recursos de apoyo en la exposición oral que sean novedosos e interactivos:

Tabla 4. Recursos y herramientas en línea para la exposición oral

Recurso	Descripción
Powtoon https://www.powtoon.com/ **VideoScribe** https://www.videoscribe.co/en/	Herramientas tecnológicas gratuitas e interactivas, mediante las cuales se pueden crear presentaciones o videos animados. Son alternativas a las presentaciones en power point y tienen el propósito de que la información no sea vista únicamente como texto plano.
Haiku Deck https://www.haikudeck.com/ **Zuru** https://www.haikudeck.com/zuru	Es una herramienta para elaborar presentaciones con mensajes directos y uso de imágenes profesionales, a fin de elaborar apoyos visuales de alto impacto que eviten la saturación de información. Además, con la extensión Zuru, esta herramienta facilita el acceso a inteligencia artificial que se encarga de realizar las presentaciones por sí sola cuidando que la información sea directa y las imágenes reflejen de la mejor forma posible el contenido de la presentación.
Zoho ShowTime https://www.youtube.com/watch?v=8dILGhys6Cw&t=52s	Es un recurso que permite al expositor interactuar con la audiencia y obtener información relevante que es útil a manera de retroalimentación sobre su ejecución y los recursos empleados durante la exposición. Así, la audiencia tiene un espacio para emitir sus valoraciones y el expositor puede recurrir a esta información para analizar su desempeño.
SlideDog https://slidedog.com/features/	Se trata de un software gratuito que permite concentrar distintos tipos de material que se utilizarán en la exposición, los ordena y presenta de forma ininterrumpida.
Google Jamboard https://www.blog.google/products/g-suite/jamboard-whiteboard-reimagined-collaboration-cloud/	Dado que en la exposición oral no solo se requiere presentar información, sino interactuar y construir conceptos en común con la audiencia, esta es una herramienta gratuita que muestra una pizarra digital interactiva que puede compartirse durante la exposición; se trata de una herramienta que forma parte de Google Suite, por lo que provee los beneficios asociados (guardado automático en la nube, compartir con otros usuarios de forma automática, entre otros).

Recurso	Descripción
Present With Impact https://covince.com/	Es una aplicación web que consiste en un aula de realidad virtual en forma de U en la que se presentan como audiencia seis humanos virtuales. Esta aplicación facilita el entrenamiento para la presentación oral y proporciona retroalimentación inmediata mediante información sobre la variación en el tono, volumen, cantidad y duración de las pausas, se puede descargar y utilizar a través de un smartphone. Fue diseñada por la empresa CoVince Adventurous learning y su uso tiene un costo que va desde paquetes básicos con un precio mensual de 8 euros, hasta paquetes premium con valor de 19 euros.

Pese a la gran cantidad de herramientas tecnológicas que existen actualmente, es importante que enfatice con sus alumnos que estos son recursos que mejoran los materiales y apoyos utilizados en la exposición oral, pero que son complementarios al resto de habilidades esperadas que están asociadas a la exposición oral. Indique que para conocer con mayor detalle las habilidades que serán valoradas en la exposición oral, deben consultar los criterios del instrumento de evaluación (rúbrica, lista de cotejo, etcétera.).

CONCLUSIONES Y REFLEXIONES FINALES

- La exposición oral es una técnica didáctica y un recurso de evaluación del y para el aprendizaje que permite el desarrollo de la comunicación oral, una de las habilidades más valoradas en el ámbito profesional en cualquier área de especialización.
- Es una técnica que implícitamente permite el desarrollo de otras habilidades como el razonamiento crítico, la capacidad de síntesis, de análisis y de argumentación, consideradas habilidades fundamentales para la vida profesional.
- Si el profesor emplea la exposición como recurso de evaluación del y para el aprendizaje en las modalidades presencial, a distancia o híbrida, debe establecer los criterios a evaluar de manera clara y precisa.
- Para que la evaluación de los alumnos a partir de esta herramienta sea útil es recomendable que ellos conozcan los criterios a evaluar en la exposición.
- Desde un marco de evaluación formativa es un recurso muy útil debido a que permite retroalimentar al alumno sobre sus fortalezas y los aspectos en los que puede mejorar.
- En modalidades educativas presenciales, a distancia o híbridas, la exposición oral permite utilizar una variedad de recursos tecnológicos de apoyo que son creativos e innovadores, con el objetivo de fomentar procesos de enseñanza y de aprendizaje más significativos y en donde el alumno es un agente activo de su aprendizaje.
- Es importante dar seguimiento antes, durante y después de la preparación de la exposición con el propósito de retroalimentar a los alumnos.

A manera de síntesis revise el video *Capítulo Exp. Oral* [Video].
Youtube. https://youtu.be/EGOhdFoJvsY

REFERENCIAS

Arce-Medina, E. (2006). La exposición oral: imperativo para los nuevos graduados. *Innovación educativa*, *6*(33), 25-32. https://www.redalyc.org/pdf/1794/179421197003.pdf

Boetje, J. y Ginkel, S. (2020). The added benefit of an extra practice session in virtual reality on the development of presentation skills: A randomized control trial. *Journal of Computer Assisted Learning, 37*(1), 253-264. https://doi.org/10.1111/jcal.12484

Cooper, L., Border, R. y Vessey, J. (2021). A Longitudinal Analysis of Student Learning Gains in Oral Competency. *Basic Communication Course Annual, 33*(10), 160-195. https://ecommons.udayton.edu/bcca/vol33/iss1/10/

Costigan, R. y Brink, K. (2020). Developing Listening and Oral Expression Skills: Pillars of Influential Oral Communication. *Journal of Management Education, 44*(2), 129-164. https://doi.org/10.1177/1052562919890895

Pimienta-Prieto, J. (2008). *Evaluación de los aprendizajes: un enfoque basado en competencias.* Pearson Education.

Teruel, M., Felipe, A., Herrera, M., Herrera, J., Guerrero, M., Dopazo, J., Díaz, M., Alzola, P. y Eyheramendy, V. (2020). Una experiencia de práctica de oralidad en la enseñanza universitaria de las ciencias morfológicas. *Ciencia, Docencia y Tecnología*, *31*(60), 286-305. https://doi.org/10.33255/3160/506

Verano, D., González, S., Bolívar, A., Fernández, M. y Galván, I. (2016). Valoración de la competencia de comunicación oral de estudiantes universitarios a través de una rúbrica fiable y válida. *Revista Brasileña de Educación*, *64*(21), 39-57. https://doi.org/10.1590/S1413-24782016216403

Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L. y Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, *28*(3), 228-260. https://doi.org/10.1080/0969594X.2021.1884042

Montoya, A., Pérez, A. y Pérez, N. [María Azucena Montoya Magno] (16 de mayo de 2022). *Capítulo Exp. Oral* [Video]. Youtube. https://youtu.be/EGOhdFoJvsY

Montoya, A., Pérez, N. y Pérez A. [Nancy Fabiola Pérez Herrera] (6 de mayo de 2022). *Exposición oral* [Video]. Youtube. https://youtu.be/oNSp4xDLVcA

Herrera, C. (2021). Evaluación para el aprendizaje: enfoque cualitativo. *Exposición oral* [MOOC] https://www.coursera.org/lecture/evaluacion-cualitativa/exposicion-XD0Kz

Capítulo 24

EXÁMENES ORALES

Carlos Gutiérrez-Cirlos, Melchor Sánchez Mendiola

"Es probable que la mayor fortaleza del examen oral es que puede usarse como un método para enseñar, más que como un instrumento de evaluación."
Patricia Houston, 2006

INTRODUCCIÓN

Las evaluaciones tipo examen oral tienen una larga historia en la educación universitaria, cuentan con amplia tradición y prácticamente todos los profesionistas las han vivido en algún momento de su vida. En Gran Bretaña las evaluaciones fueron orales hasta el siglo XIX, cuando se incluyeron los exámenes escritos, mientras que en Estados Unidos de América cuando se iniciaron pruebas escritas en Boston en 1845, se suscitaron muchas controversias. Se decía que los exámenes escritos eran superiores a las evaluaciones orales por su objetividad, imparcialidad, capacidad de cuantificación, además de la economía al realizar pruebas en grupos grandes de estudiantes. Horace Mann afirmaba que tenían las siguientes ventajas sobre las pruebas orales: se administran las mismas preguntas a estudiantes de todas las escuelas para evaluaciones imparciales; los exámenes escritos son más justos y menos estresantes en tiempo que los orales; permiten explorar un rango más amplio de preguntas; los examinadores no pueden interrumpir o guiar las respuestas de los estudiantes; se elimina el favoritismo; es más sencillo evaluar e interpretar los resultados de la prueba; y, como un daguerrotipo, muestran una imagen casi fotográfica de lo que ocurre en la mente de los estudiantes (Hanson, 1993; National Education Association, 2020).

Así que, si bien los exámenes orales fueron la norma hasta el siglo XIX, han sido sustituidos por otras formas de evaluación que buscan ser más objetivas y estandarizadas, entre otras características que han provocado que los exámenes orales se utilicen menos en la actualidad, aunque continúan siendo la herramienta de evaluación preferida para la defensa de una tesis de posgrado (maestría, candidatura al grado de doctor, doctorado y postdoctorado) y de algunas especializaciones (como en medicina) en varios lugares del mundo (Davis y Karunathilake, 2005; Memon et al., 2010). Un equivalente, sin ser de altas consecuencias,

es la evaluación de exposiciones orales realizadas por estudiantes a cualquier nivel (capítulo 23 de este libro).

La evaluación de las presentaciones orales (exposiciones, disertaciones o conferencias), incluye que el docente planee, organice y establezca los criterios a considerar en la evaluación de manera precisa, con énfasis en el adecuado conocimiento del tema y el manejo del lenguaje verbal y no verbal: el primero se refiere a explicar de manera clara, fluida, coherente, con conocimiento y convincente, un tema, mientras que el segundo implica que el estudiante adquiere confianza y seguridad en sí mismo. La evaluación de las presentaciones orales, bien realizada, puede ser de utilidad para preparar a los estudiantes para exámenes orales de altas consecuencias.

En este capítulo analizaremos las características generales de los exámenes orales, cómo pueden ser elaborados, las ventajas y desventajas que presentan, particularmente cuando son de altas consecuencias, y finalmente, algunos ejemplos de su uso en los estudios de pregrado y posgrado.

¿QUÉ SON LOS EXÁMENES ORALES?

El examen oral o *viva voce* (de viva voz), es un método de evaluación en el que uno o más evaluadores hacen preguntas de forma verbal al estudiante, quien, a su vez, también responde de forma oral. En este diálogo, es posible evaluar competencias que en otro tipo de evaluaciones es más complicado, tales como: profundidad del conocimiento, su aplicación, organización de conceptos de manera rápida y esquemática, evaluación del razonamiento y criterio, entre otros (Pérez, 2015). Es un método conocido por la mayoría de los estudiantes y docentes y fue la forma predominante de evaluar hasta el siglo XIX. Existen varios tipos de exámenes orales, que se resumen en la siguiente tabla (Tabla 1):

Tabla 1. Tipos de exámenes orales

Tipo de examen	Ejemplo
Examen oral tradicional no estandarizado	Evaluar aspectos decididos por los docentes con un caso seleccionado al azar, tiempo variable.
Examen oral estandarizado	Evaluar un caso con criterios de grupo, se ponderan las respuestas, tiempo definido.
Examen con un caso corto	En matemáticas: explique las aplicaciones de la derivada. En medicina: explique el manejo de un caso de diabetes descompensada.
Examen con un caso largo	En una escuela de negocios: presente una estrategia para convencer a sus clientes.
Examen de grado	Proceso para la certificación como especialista de anestesiología.
Defensa de tesis de posgrado	Ante un jurado conformado por presidente, secretario y vocales, hacer una presentación oral y defender la tesis.

En los exámenes orales puede haber uno o varios examinadores, y pueden aplicarse a uno o varios estudiantes. Estas pruebas se pueden llevar a cabo con fines formativos, sumativos o como exámenes de altas consecuencias, como son los exámenes de grado de licenciatura o de especialidad o la defensa de una tesis de posgrado.

¿CÓMO ELABORAR UN EXAMEN ORAL?

Existen algunas referencias recientes sobre cómo elaborar de manera adecuada un examen oral, aunque mucho depende de las reglas generales que cada institución educativa tiene para su aplicación y de los usos y costumbres del profesorado. El estilo y personalidad de cada examinador puede variar y es claro que su conducta es crucial en este tipo de proceso de evaluación. Es posible hacer algunas recomendaciones generales para que la evaluación sea en lo posible, lo más justa para los estudiantes (Memon et al., 2010; Mohammed, 2001): la actitud de los evaluadores debe ser lo más amigable posible, generando un espacio psicológicamente seguro para el o la sustentante; el objetivo del examen debe estar bien definido, ya sea para evaluar competencias generales, desempeño o defensa de una tesis de investigación; el puntaje obtenido debe basarse en el uso de un instrumento estructurado con evidencia de validez y confiabilidad, como una lista de cotejo o rúbrica; cada evaluador idealmente debe generar su puntaje individual antes de dar el puntaje global final con los demás jueces.

Considere que los estudiantes pueden estar muy ansiosos durante los exámenes orales y lo anterior puede correlacionar con evaluaciones menos favorables (Thomas, 1993). Los evaluadores deben mostrar respeto y, en lo posible, saludar al estudiante, presentándose por su nombre. Esto puede facilitar que el estudiante se sienta menos ansioso y con más confianza, lo que no debe implicar que los examinadores sean menos exigentes. Dicha conducta se debe aplicar durante todo el tiempo que dure la evaluación y con los diferentes estudiantes evaluados.

Durante la fase inicial, el evaluador debe permitir que el examinado se exprese libremente, sin interrupciones, lo que mejora el desempeño y disminuye la ansiedad en los siguientes minutos. Se debe establecer el tiempo de la presentación, y al concluir, iniciar con las preguntas. Es aconsejable comenzar con preguntas más sencillas y generales para después realizar preguntas más complejas y específicas, con una pregunta a la vez. Mantenga contacto visual con el evaluado, no se distraiga con dispositivos electrónicos (podría parecer que está aburrido o no totalmente involucrado en el proceso) y muestre un lenguaje corporal, verbal y facial lo más ecuánime posible. Trate de no discutir la respuesta con el evaluado, sobre todo antes de que la respuesta concluya. En lo posible, como evaluador, compórtese con seriedad y evite las bromas, lo que no debe impedir que se muestre paciente y amigable, sobre todo al final de la discusión o al final del día, cuando un examen ha sido más largo de lo esperado o ha aplicado varios exámenes. Las palabras sobre el desempeño deben darse después de la discusión con el grupo de evaluadores, no antes, para no generar falsas expectativas en el estudiante.

Las personas con mejores habilidades para comunicarse pueden ser mejor evaluadas. No se guíe por el sexo, edad, aspecto, arreglo personal, estatura, origen étnico, entre muchos aspectos que pueden introducir sesgos y hacer injusta la evaluación. No se guíe por el "beneficio de la duda" para aprobar a un estudiante: su objetivo no es ayudar al estudiante a aprobar el examen, sino ayudarle a aprender y mejorar su desempeño. Por otra parte, la comunidad requiere tener profesionistas competentes, por lo que las evaluaciones orales sumativas deben tener el rigor necesario. La evaluación debe ser independiente y discutida con el grupo hasta que todos terminen de dar su puntaje, lo que puede evitar que suba o baje el puntaje por influencia de los otros (Schwiebert, 1993).

Ventajas de los exámenes orales

- Al examen oral se le han atribuido una serie de ventajas y potencialidades, como son:
- la familiaridad de la comunidad tanto de docentes como estudiantes con el formato, lo cual lo hace aceptable;
- este método utiliza el medio de comunicación natural de los seres humanos, el habla, lo que facilita su aceptación y da la impresión que no se requiere mayor entrenamiento para realizarlo;
- promueve el contacto personal directo entre las personas, lo que incrementa su autenticidad comparado con un examen escrito en donde el reto es enfrentarse a un documento estático en papel o en línea que generalmente tiene solo texto o imágenes;
- se dice que con el método se evalúan habilidades cognitivas superiores, como la solución de problemas, el pensamiento crítico y las habilidades de razonamiento;
- proporciona la oportunidad de explorar el conocimiento con mayor profundidad, ya que el o los evaluadores están enfocados en el desempeño del estudiante individual;
- tiene flexibilidad para que el examinador explore diversos caminos y contenidos, lo que depende del tipo y calidad de respuestas del sustentante;
- tiene mayor posibilidad de retroalimentación inmediata y directa, ya que el proceso ocurre de forma sincrónica, en tiempo real;
- también se puede adecuar el examen a las necesidades del estudiante individual.

Muchas de estas virtudes no han sido demostradas con trabajos de investigación rigurosos, incluso algunas de ellas han sido cuestionadas y refutadas por varios académicos, por lo que hay que ser prudente en su interpretación.

Sin duda, otra de las ventajas de los exámenes orales es que pueden evaluar los cuatro niveles propuestos por George Miller: saber, saber cómo, mostrar cómo y realizar la actividad en escenarios reales (Miller, 1990). Inclusive se ha propuesto un nivel superior, en el que además se puede evaluar si el desempeño de la actividad se puede delegar en una persona con niveles mínimos de supervisión o sin supervisión (ten Cate, 2020). Un ejemplo en el área de ciencias de la salud, consiste en la evaluación de las actividades clínicas a confiar o del razonamiento clínico, como niveles agregados a evaluar además de realizar la actividad, que pueden ser documentados con un examen oral (Witheridge et al., 2019). Lo anterior

se ilustra en la Figura 1, con un ejemplo de evaluación oral para descartar el diagnóstico de una enfermedad del corazón como el infarto. Un examen oral bien diseñado y aplicado para este caso, podría medir la combinación de desenlaces relacionados con lo verbal (saber qué), intelectual (saber cómo), y habilidades cognitivas (conocimiento estratégico, metacognición, razonamiento clínico, actividades clínicas confiables).

Figura 1

También es importante mencionar que el examen oral puede llevarse a cabo en espacios formales e informales de aprendizaje, en diversos momentos y tiempos, lo cual le proporciona cierta flexibilidad. Lo anterior implica que docente y estudiante establezcan un compromiso de lo que se evaluará, con adaptabilidad del espacio y del tiempo para las dos partes, además, si se toma en cuenta el nivel de conocimiento del estudiante se puede dar una retroalimentación adecuada, lo que puede no suceder con otro tipo de exámenes, como en los de opción múltiple aplicados a grandes grupos, en los que es difícil la retroalimentación por múltiples causas: elevado número de estudiantes, pocos profesores, seguridad de la base de datos, entre otras.

Desventajas de los exámenes orales

Todos los que hemos vivido exámenes orales, sobre todo los de índole sumativa, hemos experimentado alguna de las siguientes emociones: miedo, tensión, ansiedad, nerviosismo, pánico escénico, incluso en ocasiones crisis de llanto o de depresión. El desempeño en un examen oral ante un juez o un jurado evaluador puede ser una experiencia amenazadora y estresante para la persona evaluada, por lo que debe llevarse a cabo con cuidado y profesionalismo. Lo anterior puede provocar que el estudiante evaluado falle en las respuestas, aunque conozca los conceptos evaluados, lo que introduce la amenaza a la validez de varianza irrelevante al constructo. Algunas de las desventajas de esta modalidad de evaluación son las siguientes:

- Puede ser difícil distinguir entre lo que el estudiante dice y cómo lo dice, separar el discurso de los conceptos vertidos en la respuesta no es trivial;
- aunque se piense lo contrario, en exámenes orales que no son de altas consecuencias, está demostrado que usualmente se usan para evaluar niveles bajos de procesos cognitivos como el simple recuerdo de datos memorizados;
- generalmente hay falta de rutas administrativas o documentación escrita para realizar revisiones o impugnaciones por parte de los sustentantes, lo que puede generar una condición de vulnerabilidad, sobre todo en las evaluaciones sumativas en las que se les informa a los estudiantes que "la decisión del jurado es inapelable", aunque no se tenga mucha claridad del proceso de obtención de las puntuaciones y de la toma de decisión de aprobar o no el examen;
- el impacto de las "primeras impresiones" que tienen los jueces de los estudiantes al inicio de la evaluación, puede tener influencia desproporcionada en la calificación, aunque se tenga un buen desempeño en el resto del examen;
- la principal limitación sobre la que existe mayor cantidad de evidencia de investigación, es su baja confiabilidad, lo que impacta fuertemente a la validez, por lo que se ha dejado de usar en exámenes sumativos en muchas partes del mundo (Davis y Karunathilake, 2005; Memon et al., 2010);
- la influencia de los sesgos del evaluador, se ha demostrado que hasta el 27% de la varianza de los resultados de un examen oral se deben al examinador;
- puede haber múltiples sesgos al tomar en cuenta de forma consciente o inconsciente, aspectos ajenos al constructo objetivo de la prueba, como aspecto físico, vestuario, fluidez, facilidad de palabra, género, color de piel, extroversión, entre otros, lo que tiene profundas implicaciones sobre todo cuando se usa con intención sumativa (Memon et al., 2010);
- este tipo de exámenes requiere una inversión sustancial de recursos y de tiempo de los examinadores, que, si son profesionales ocupados, se convierte en un reto para que la evaluación pueda completarse en tiempo y de forma adecuada.

Es pertinente tomar en cuenta el fenómeno de "especificidad de caso", ya que en el formato oral se limita frecuentemente a evaluar solo algunos problemas o casos, y está demostrado

que el desempeño de una persona en un caso no predice el desempeño en otros casos y contextos. No se puede generalizar que, si un estudiante tiene un alto desempeño en un examen oral con un caso relativamente sencillo, tendrá el mismo desempeño con casos diferentes, lo que puede disminuir la confiabilidad de la prueba. Como ejemplo, en la Facultad de Medicina de la Universidad Nacional Autónoma de México (UNAM), el estudiante podía graduarse como médico mediante un examen oral ante paciente real (APR). El estudiante era citado en un hospital y un jurado de al menos cuatro sinodales, lo evaluaba ante un paciente que un día antes había aceptado ser parte del examen. Podía ser un paciente fácil o difícil, lo que dependía del hospital, del jurado y de la disponibilidad de pacientes. Si bien se solicitaba a los sinodales que se eligiera un paciente relacionado con la medicina general, lo anterior no sucedía siempre y el estudiante se enfrentaba a pacientes con enfermedades raras, no llegaba al diagnóstico y esto podía interpretarse como falta de conocimiento y mal desempeño. Esto es un ejemplo de la "especificidad de caso": si el estudiante se hubiera enfrentado a un paciente con un diagnóstico menos complejo o si hubiera tenido la fortuna de haber estudiado precisamente ese caso la noche previa al examen, su desempeño hubiera sido diferente. Por lo anterior, entre otros factores, actualmente en la Facultad de Medicina de la UNAM, se utilizan otras modalidades al momento de graduarse, como el Examen Clínico Objetivo Estructurado (ECOE) junto con un examen general de conocimientos y el examen APR es realizado por un bajo porcentaje de los estudiantes (Ponce de León et al., 2006).

En los exámenes orales el bajo muestreo de contenido puede ser una limitación grave, ya que generalmente el tiempo de examen no es el suficiente para explorar todos los contenidos indicados en el currículo, y en consecuencia implica realizar juicios de valor con evidencia de validez de contenido limitada. Puede haber falta de estandarización y tendencia a improvisar durante el examen, dependiendo del tiempo y estados de ánimo de los jueces, lo que impacta la confiabilidad del examen, ya que cada estudiante se sujeta a un examen diferente. También hay que tomar en cuenta la faceta cultural del currículo oculto de cada profesión, como dijo un colega médico: *"con los exámenes orales no medimos gran cosa, solamente juzgamos si el estudiante es apto o no para unirse al club de profesionistas registrados"* (Marshal y Ludbrock, 1972).

Aun cuando se busca controlar los factores mencionados, la personalidad y la ansiedad influyen en los resultados de la evaluación oral, hecho documentado en un estudio realizado sobre el examen oral final de un curso de psiquiatría a 40 estudiantes de medicina, con seis evaluadores expertos (Thomas, 1993). La forma en que el sujeto responde (confianza en sí mismo), correlacionó con evaluaciones positivas, mientras que la presencia de ansiedad, particularmente en hombres, con evaluaciones negativas. El sexo, atractivo físico, atributos personales (extraversión, inestabilidad emocional) fueron poco determinantes en los resultados del examen oral, sin embargo, cuando los estudiantes fueron evaluados por los investigadores fueron reprobados, mientras que los evaluadores expertos los habían pasado a todos, lo que habla de la baja reproducibilidad de este tipo de evaluación (Thomas, 1993). Los autores de este estudio comentaron que, para eliminar posibles injusticias, el examen se realizaba con dos evaluadores, y encontraron que el 100% de los residentes eran aprobados.

Lo que ocurría es que, ante el desacuerdo, acordaban aprobar al residente y no reprobarlo, fenómeno frecuente en evaluación en educación superior conocido como “fracaso para fallar” (*failure to fail*, en inglés) (Yepes-Rios et al., 2016).

¿Es el examen oral el estándar de oro o el oro de los tontos? (Houston, 2006). La reproducibilidad, relacionada con la precisión y consistencia de un examen, en el caso de las evaluaciones orales depende de la variabilidad relacionada con: el examinador, el examen, y el sustentante. Para la certificación de especialistas en anestesiología por el Consejo Americano de Anestesiología (ABA, por sus siglas en inglés), se ha utilizado desde hace más de 70 años el examen oral (Jacobsohn, 2006). Al evaluar a 25 médicos residentes con un examen simulado parecido al del ABA, con dos examinadores (E1 y E2) con seis semanas de diferencia, la correlación entre los evaluadores fue pobre: para los E1 de 0.243 y para los E2 de 0.405. Para 48% de los examinados, la probabilidad de pasar o no, dependió en un 48% del evaluador. Lo anterior mejora cuando más exámenes orales se realizan, con reproducibilidad de hasta 0.79 y se ha demostrado que cuando el desempeño del evaluado es pobre, los evaluadores son menos consistentes en el resultado. Al parecer, lo que mejora el desempeño en los exámenes orales, es hacerlos, en este caso, desde el inicio de la residencia. Dar una clase sobre el tema, preparar de forma estandarizada a los evaluadores, un nivel de competencia similar de los examinados, actividades estandarizadas, parecen ser menos determinantes que las respuestas no verbales observadas por los examinadores. Por lo anterior, la competencia debe medirse con diversos métodos de evaluación: orales, escritos y de desempeño, en vivo o con simuladores. La preparación para un examen oral puede, finalmente, darle más habilidades al evaluado para manejar de mejor manera el conocimiento requerido, además de desarrollar habilidades para integrar y comunicar ideas relacionadas con la práctica reflexiva.

EJEMPLOS DEL USO DE EXÁMENES ORALES

a) Examen oral (formativo) en línea para revitalizar las clases en línea

Como consecuencia de la migración a la educación en línea durante la pandemia por COVID-19, se publicaron estudios sobre el uso del examen oral en las condiciones de educación remota de emergencia, con buenos resultados (Dumbaugh, 2020; Sabin et al., 2021). Se encontró que realizar evaluaciones en línea puede mejorar el tiempo para la evaluación, aumentar la seguridad para los estudiantes y también ayuda a mantener el interés en las clases en línea, lo que es un reto persistente en muchas instituciones como consecuencia de la pandemia. Con exámenes orales sobre temas breves (por ejemplo: describa la utilidad de la derivada) y eventos programados, mejora la confianza de los estudiantes, los ayudan a disminuir la ansiedad, aumentan la creatividad para presentar y dar soluciones, además de que se otorga retroalimentación de forma inmediata. Los exámenes son grabados, lo que permite a los estudiantes revisar la solución de los problemas paso a paso. La satisfacción para los estudiantes y el profesorado es mutua y consideran que estos “exámenes orales breves” en línea originados por la pandemia, se convertirán en parte de su proceso didáctico.

b) Examen oral (sumativo) en una escuela de negocios

Son múltiples las habilidades que un estudiante de negocios y administración debe desarrollar y que pueden ser evaluadas y mejoradas en un examen oral: dar un discurso ante un auditorio, organizar sus ideas para informar y convencer a la audiencia, saber citar información correcta para sostener sus afirmaciones, demostrar habilidades de comunicación para escuchar de forma comprensiva, manejar conflictos interpersonales y de grupo, entender las habilidades, roles y métodos para aumentar la motivación, productividad y satisfacción para obtener desenlaces de alta calidad y finalmente entender que la comunicación es el instrumento para tener un intercambio eficiente de ideas y de información (Markulis, 2008). Lo anterior puede lograrse mediante varias formas, pero el examen oral puede ser una herramienta que permite evaluar los aspectos mencionados.

c) Examen oral (de altas consecuencias) como defensa de tesis de grado

La pandemia por COVID-19, provocó que la educación presencial cara-a-cara prácticamente cesara a todos los niveles, incluyendo el posgrado (Gupta et al., 2020). Lo anterior no implicó que se haya detenido la impartición de clases, asistencia a cursos y evaluación de los estudiantes, ya que estas actividades se llevaron a cabo en diversas plataformas en línea. En muchas instituciones de educación superior se implementó el examen de grado a distancia por medio de las plataformas digitales de videoconferencia. El estudiante envía su tesis al jurado y programa el examen en modalidad en línea, sincrónico, en la que las sesiones pueden ser grabadas. Estos exámenes se han estado realizando en diversas universidades a nivel global con buenos resultados.

En épocas recientes se ha explorado su pertinencia en evaluación en línea y a distancia, ya que el examen oral tiene ventajas en este contexto: se puede grabar con facilidad y utilizar el video para evaluaciones formativas o sumativas; se puede dar retroalimentación inmediata; el entorno virtual es menos jerárquico y puede disminuir la sensación de ansiedad y estrés; por supuesto que hay necesidad de infraestructura tecnológica y habilidades en su uso, aunque actualmente ha disminuido este reto a nivel universitario; tanto estudiantes como examinadores pueden estar en lugares distintos geográficamente, lo cual es una gran ventaja; y se puede hacer uso de múltiples recursos tecnológicos y de aplicaciones informáticas para la presentación, incluyendo medidas para mejorar la accesibilidad visual o auditiva. En muchas universidades se ha llevado a cabo con éxito esta modalidad, incluso para exámenes finales o de obtención de grado, como en la Universidad Nacional Autónoma de México.

CONCLUSIONES

A pesar de sus limitaciones, el examen oral continúa siendo una estrategia muy utilizada en todo el mundo, tanto con intencionalidad formativa como sumativa y de alto impacto. Es previsible que seguiremos aplicando exámenes orales en el futuro, lo que debe comprometernos a realizarlos lo mejor posible, siguiendo las siguientes recomendaciones:

- Usar los principios fundamentales de una buena evaluación (validez, confiabilidad, justicia, aceptabilidad, etc.), que son el pilar fundamental del uso de cualquier instrumento de evaluación.
- No debemos usarlos para evaluaciones que se puedan hacer mejor con otros instrumentos (por ejemplo, para la evaluación de conocimiento, que se logra mejor con exámenes escritos), hay que usarlo para lo que realmente sirva y sea una mejor opción.
- Es importante estructurar el proceso y las preguntas, esto mejora la confiabilidad.
- Es recomendable usar varios exámenes y examinadores.
- Procurar hacer las mismas preguntas o similares a los estudiantes.
- Usar descriptores, rúbricas y criterios específicos para las respuestas.
- Si es necesario videograbar el examen, hay que hacerlo cuidando los aspectos éticos y legales, además de los técnicos.
- Entrenar a los examinadores es fundamental. Los exámenes orales son actualmente un instrumento más a utilizar en las estrategias de evaluación, como complemento del resto de las herramientas, usémoslos con prudencia y sensatez, procurando en todo momento mejorar el aprendizaje de los estudiantes.

REFERENCIAS

Davis, M. H., & Karunathilake, I. (2005). The place of the oral examination in today's assessment systems. *Medical Teacher, 27*(4), 294–297. https://doi.org/10.1080/01421590500126437

Dumbaugh, D. (2020, 9 septiembre). How to use oral examinations to revitalize online classes (opinion). https://www.insidehighered.com/advice/2020/09/09/how-use-oral-examinations-revitalize-online-classes-opinion

Gupta, V. S., Kapur, M., Naik, M., & Mohammad, A. (2020). Post-graduate exams amidst COVID-19 pandemic: Our experience. *Indian journal of ophthalmology*, *68*(11), 2630–2631. https://doi.org/10.4103/ijo.IJO_2672_20

Hanson, F. A. (1993). Testing Testing: Social Consequences of the Examined Life.

Berkeley: University of California Press. http://ark.cdlib.org/ark:/13030/ft4m3nb2h2/

Houston, P., Kearney, R. A., & Savoldelli, G. (2006). The oral examination process - gold standard or fool's gold. *Canadian Journal of Anaesthesia = Journal canadien d'anesthesie*, *53*(7), 639–642. https://doi.org/10.1007/BF03021620

Jacobsohn, E., Klock, P. A., & Avidan, M. (2006). Poor inter-rater reliability on mock anesthesia oral examinations. Canadian Journal of Anaesthesia = *Journal canadien d'anesthesie, 53*(7), 659–668. https://doi.org/10.1007/BF03021624

Markulis, P. M., & Geneso, S. (2008). «Viva Voce»: Oral exams as a teaching & learning experience. Developments in Business Simulation and Experiential Learning, 35, 118–127.

Marshall, V. R., & Ludbrook, J. (1972). The relative importance of patient and examiner variability in a test of clinical skills. *British Journal of Medical Education*, *6*(3), 212–217. https://doi.org/10.1111/j.1365-2923.1972.tb01660.x

Memon, M. A., Joughin, G. R., & Memon, B. (2010). Oral assessment and postgraduate medical examinations: establishing conditions for validity, reliability and fairness. *Advances in health sciences education: theory and practice*, *15*(2), 277–289. https://doi.org/10.1007/s10459-008-9111-9

Miller G. E. (1990). The assessment of clinical skills/competence/performance.
Academic medicine: journal of the Association of American Medical Colleges, *65*(9 Suppl), S63–S67. https://doi.org/10.1097/00001888-199009000-00045

Mohammed, MSJ., Al Buhairi ,AR. (2000). Guidelines for the Administration of Oral Examinations: Review and Opinion. *Saudi Medical Journal*, *21*(11), 1013-1015.

Morgado Bernal, I. (2019, 6 marzo). En defensa del examen oral. *El país*. https://elpais.com/elpais/2019/02/18/ciencia/1550486209_570944.html

National Education Association. (2020). History of Standardized Testing in the United States. 9 de octubre 2021, *National Education Association* Sitio web: https://www.nea.org/professional-excellence/student-engagement/tools-tips/history-standardized-testing-united-states

Ponce de León-Castañeda, M. E., y Ortiz-Montalvo, A., y Petra-Micu, I., y Ortega-Gómez, J. L. (2006). Fase práctica del examen profesional: validación de un cuestionario de opinión de alumnos. *Archivos en Medicina Familiar, 8*(3),183-189. Disponible en: https://www.redalyc.org/articulo.oa?id=50780307

Sabin, M., Jin, K.H., & Smith, A. (2021). Oral Exams in Shift to Remote Learning. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education* (SIGCSE '21). Association for Computing Machinery, New York, NY, USA, 666–672. DOI: https://doi.org/10.1145/3408877.3432511

ten Cate, O., Carraccio, C., Damodaran, A., Gofton, W., Hamstra, S. J., Hart, D. E., Richardson, D., Ross, S., Schultz, K., Warm, E. J. Whelan, A. J. & Schumacher, D. J. (2021). Entrustment Decision Making: Extending Miller's Pyramid. *Academic Medicine*, *96*(2), 199-204. https://doi.org/10.1097/ACM.0000000000003800

Pérez H, I., Vergara R, C., Goens G C., Viviani G, P., Letelier S, L. (2015). Percepción de examen oral estandarizado vs no estandarizado en el internado de medicina interna. *Revista Médica de Chile*, 143, 841-846.

Schwiebert P., Davis A. (1993). Increasing inter-rater agreement on a family medicine clerkship oral examination–a pilot study. *Family Medicine. 25*(3):182-185.

Thomas, C. S., Mellsop, G., Callender, K., Crawshaw, J., Ellis, P. M., Hall, A., MacDonald, J., Silfverskiold, P., & Romans-Clarkson, S. (1993). The oral examination: a study of academic and non-academic factors. *Medical Education*, *27*(5), 433–439. https://doi.org/10.1111/j.1365-2923.1993.tb00297.x

Witheridge, A., Ferns, G., & Scott-Smith, W. (2019). Revisiting Miller's pyramid in medical education: the gap between traditional assessment and diagnostic reasoning. *International Journal of Medical Education*, *10*, 191–192. https://doi.org/10.5116/ijme.5d9b.0c37

Yepes-Rios, M., Dudek, N., Duboyce, R., Curtis, J., Allard, R. J., & Varpio, L. (2016). The failure to fail underperforming trainees in health professions education: A BEME systematic review: BEME Guide No. 42. *Medical Teacher*, *38*(11), 1092–1099. https://doi.org/10.1080/0142159X.2016.1215414

Capítulo 25

SIMULACIÓN

Guadalupe Soto Estrada, Laura S. Hernández Gutiérrez

"Dime y lo olvido, enséñame y lo recuerdo, involúcrame y lo aprendo."
Benjamín Franklin

INTRODUCCIÓN

La simulación es una técnica didáctica empleada en las instituciones de educación superior que se incorporó en el ámbito de la educación a mediados del siglo XX y surgió dentro del espacio educativo que interrelaciona juegos y simulaciones. Gaba (1988) la define como una técnica, no una tecnología, para sustituir o ampliar las experiencias reales a través de experiencias guiadas, que evocan o replican aspectos sustanciales del mundo real, de una forma totalmente interactiva.

El objetivo de la simulación es la enseñanza, el aprendizaje y la evaluación de los estudiantes, situado en un contexto que imite el entorno que se quiera simular de manera segura. Cuando el profesor está consciente del papel activo que debe jugar el estudiante durante su proceso de aprendizaje, de lo importante que es guiar, realimentar y evaluar de forma continua este proceso, es inevitable que implemente en su práctica pedagógica una evaluación cuyo objetivo sea obtener constantemente evidencias del progreso e identificar las áreas de oportunidad de los estudiantes para potenciar al máximo sus capacidades por lo que la simulación es una actividad que contribuye en gran medida en la evaluación de los estudiantes.

De esta forma, la técnica permite que los estudiantes apliquen sus conocimientos y habilidades en una situación parecida a la realidad mediante escenarios controlados que favorecen el desarrollo de capacidades y destrezas indispensables en la formación profesional, como es la capacidad para argumentar, analizar, razonar, reflexionar, tomar decisiones y resolver problemas, entre otras. Es importante destacar que si bien la simulación *per se*, proporciona información muy valiosa y rica sobre el proceso de aprendizaje de los estudiantes, también es un recurso para la evaluación.

Además, si se complementa con otras metodologías activas como el Aprendizaje Basado en Problemas (ABP), por ejemplo, empleando para su evaluación instrumentos como la lista

de cotejo y la rúbrica, resulta en un mayor beneficio para el aprendizaje y la evaluación de los estudiantes, con posibilidad de implementarse de forma presencial y a distancia.

Este texto está enfocado en explicar de manera práctica en que consiste la simulación, desde su diseño, aplicación, análisis de resultados, así como la utilización de esta técnica a distancia con la finalidad de que el profesor valore la utilidad que tiene en el fortalecimiento del aprendizaje, además de ser muy útil para evaluar el grado de avance del alumno tanto a nivel individual como grupal.

¿QUÉ ES?

La simulación es considerada actualmente una valiosa técnica de evaluación que consiste en valorar el desempeño de los estudiantes o profesionales en escenarios controlados de distintas disciplinas. El escenario forma la base del aprendizaje basado en la experiencia, y es donde se lleva a cabo una representación que imita lo más cercano a una realidad, que sitúa a los alumnos en un contexto en el que se establecen situaciones o problemas similares a los que deberán enfrentar en la vida o durante su desarrollo profesional. Además, permite que el alumno pueda observar procesos complejos y analizar cambios al modificar variables o secuencias.

Su utilización es más conocida en la educación media superior y superior en áreas como las Ciencias fisicomatemáticas, las Ingenierías, las Ciencias biológicas, químicas y de la salud, las Ciencias sociales y las Humanidades. Sin embargo, se puede emplear en todos los niveles educativos y áreas del conocimiento.

TIPOS DE SIMULACIONES

Existen diversas clasificaciones en la literatura en cuanto a las alternativas de simulación basadas en el concepto de fidelidad este término define el grado de realismo de los modelos utilizados.

Simulación de baja fidelidad

Promueve el desarrollo de habilidades psicomotrices que centran su importancia en la habilidad y la destreza. Para ello, se utilizan entrenadores de tareas por partes (*part task trainers*), que requiere espacios donde se dispone de un simulador para el entrenamiento sistematizado de un procedimiento, estas habilidades se realizan mediante práctica deliberada. Aquí, la evaluación está basada en el logro del estudiante y se lleva a cabo durante o inmediatamente después de la práctica, a través de la observación, listas de cotejo o rúbricas, donde la retroalimentación oportuna, cobra relevancia para la evaluación formativa.

Simulación de mediana fidelidad

Busca la demostración de competencias de predominio técnico y la interacción con una persona con la finalidad de hacer evidentes algunas competencias profesionales, actitudinales,

comunicativas, de interés y, al mismo tiempo, poder ver la aplicación de habilidades y destrezas. Por las características de las competencias a lograr, la simulación de mediana fidelidad se puede desarrollar en escenarios de simulación con objetivos de aprendizaje bien definidos.

Simulación de alta fidelidad

Ayuda a evidenciar el desempeño de los alumnos, así como su actitud, responsabilidad, toma de decisiones, liderazgo, trabajo en equipo, entre otras, en un ambiente seguro de aprendizaje, lo que implica la posibilidad de equivocarse y aprender de la equivocación o del acierto dentro de un ambiente muy cercano a la realidad. En este tipo de simulación los escenarios son fundamentales donde el diálogo guiado por el instructor (*debriefing*) favorece el análisis y la reflexión de lo ocurrido durante el escenario.

En los escenarios diseñados y el uso de los simuladores de alta fidelidad, se pueden modificar los valores de las variables y parámetros, con ello probar diferentes hipótesis y condiciones sobre un problema, lo que permite evaluar en el estudiante las destrezas adquiridas, antes de enfrentarse a situaciones reales.

El desarrollo de las nuevas tecnologías, que pueden ser aplicadas a la simulación, ha impulsado la creación de centros de simulación en diversas áreas del conocimiento.

Actualmente, la simulación está incrementando su prevalencia y diversas instituciones la han adoptado como herramienta de evaluación que contribuye, mediante técnicas innovadoras, al aprendizaje de los estudiantes en diversas áreas del conocimiento.

Ventajas

- La simulación, como estrategia de evaluación, resulta muy útil cuando los sujetos en los que se va a intervenir o con los que se va a interrelacionar el estudiante en la vida real no están disponibles, o cuando es incorrecto probar diferentes soluciones alternativas en un sujeto real, o bien, cuando es importante desarrollar habilidades sensoperceptuales o intelectuales previas a su trabajo como profesionales.
- Permite acelerar el proceso enseñanza-aprendizaje, posibilita la contextualización de los contenidos revisados en el espacio educativo.
- Evita accidentes o eventos adversos, cuando se llevan a cabo actividades de experimentación por parte de los estudiantes.
- Permite a los docentes proporcionar una retroalimentación inmediata, además de fomentar el aprendizaje reflexivo.
- Hoy en día existe evidencia suficiente de que la utilización de la simulación en la enseñanza permite aumentar el rendimiento académico, la motivación, así como la mejora de las habilidades prácticas al mismo tiempo que permita descartar conceptos equivocados o mal interpretados por parte de los estudiantes.

Limitaciones

- Aunque la simulación como herramienta de evaluación es muy útil, su diseño es complejo ya que requiere del estudiante diversas competencias que van desde el razonamiento lógico y crítico, hasta la capacidad para resolver problemas, así como para tomar decisiones y trabajar en equipo.

¿CÓMO LA DISEÑO?

La simulación, como herramienta de evaluación, requiere de una definición precisa de la situación a la que se va a enfrentar el estudiante. Su diseño depende de los resultados de aprendizaje que se quieran observar en los estudiantes, a partir de esto se eligen las actividades por simulación necesarias ya sea una práctica deliberada, o un escenario. Una vez definidos los resultados de aprendizaje y las actividades por simulación se planean y organizan los recursos necesarios para la actividad, ya sean tecnológicos, como maniquíes o computadoras, o humanos como los pacientes simulados (utilizados para formación médica).

Uno de los modelos de simulación más conocidos es el denominado EAC (Entornos de aprendizaje constructivista) que consiste en que se parte de un problema, pregunta o proyecto que el alumno debe resolver y que difiere de los modelos objetivistas en los que se parte de conceptos y de información a partir de los cuáles se desarrolla la simulación.

La evaluación por medio de la simulación requiere de varios pasos a seguir:

1) Analizar las condiciones (recursos humanos y materiales) que existen para la evaluación mediante la simulación.
2) Establecer los propósitos de la evaluación de acuerdo con los contenidos del programa académico establecido (mapa curricular, perfil de egreso y programa de la asignatura). Se recomienda seleccionar los aprendizajes que se esperan de los estudiantes y que al mismo tiempo sirvan como criterios para evaluar. Se requieren objetivos congruentes con el marco curricular y con los componentes del perfil del egresado. Por lo tanto, la simulación funciona mucho mejor cuando está integrada al currículo.
3) Definir si el tipo de evaluación será diagnóstica, formativa o sumativa. Esto depende del momento en que se realice la simulación, así como la finalidad que se tiene respecto al resultado obtenido.
4) Definir el tipo de simulación, ya sea practica deliberada o escenarios de simulación, así como los elementos a utilizar ya sean reales, físicos, documentados o virtuales y las condiciones en las que se presentarán. Por lo tanto, en esta etapa se definen los aspectos del aprendizaje que se desean valorar (conocimientos, habilidades, actitudes).
5) Diseño de la práctica, por lo que es muy recomendable que el diseño de la evaluación con simulación se realice por un experto en la técnica para su diseño, en colaboración con expertos en el área a evaluar que aporten ideas para su diseño y desarrollo. Este punto es de gran relevancia porque aquí se planea toda la actividad por

simulación, se desarrollan los casos problemas para los escenarios y los pasos para la práctica deliberada.

6) En ocasiones, debido al empleo de material de apoyo de alta sofisticación, se pueden requerir pruebas previas de los escenarios, de los simuladores para asegurar su funcionamiento, por lo que es frecuente que este tipo de escenarios adquiera un perfil institucional.
7) Posteriormente, se determinan los criterios de evaluación con base en el aprendizaje esperado y las características del ejercicio de simulación. Es recomendable elegir el instrumento que apoyará la evaluación, que puede ser: lista de cotejo, rúbrica, guía de observación o escala de calificación.
8) Se establecen las características que tendrá la simulación, así como el papel tanto del profesor como de los alumnos. Esto con base en los objetivos planteados anteriormente (asignar los roles, las responsabilidades, las participaciones y las condiciones en el escenario simulado).
9) Se organizan los grupos de trabajo y se asignan los papeles que desempeñará cada miembro, ya sea que se realice de forma individual o grupal. Todos los integrantes deben haber entendido qué van a hacer y con qué instrumentos cuentan para analizar y abordar la situación que enfrentan, así como establecer la solución que cada grupo considere más adecuada.
10) Se informa sobre los propósitos de la evaluación, así como los criterios e instrumentos que serán utilizados mediante una sesión de *prebriefing* o sesión informativa previa, donde además se deberán firmar contratos de realidad y de ficción por parte de los estudiantes, así como la autorización de una videograbación de la práctica con fines educativos y de investigación.

Recomendación:

Un criterio para la elección de simuladores, que sirva tanto para el aprendizaje como para la evaluación, será aquella que cuente con los elementos clave de la competencia y refleje la complejidad de la realidad. Si se incorporan elementos de reto e innovación, al estilo de los videojuegos, las simulaciones resultan más interesantes y útiles para el aprendizaje.

¿CÓMO LA APLICO?

Una vez realizada la simulación, se lleva a cabo la evaluación mediante el *debriefing* o diálogo análisis-reflexión, lo que permite que los participantes comenten en primer lugar las emociones durante la práctica (ansiedad, miedo, enojo, frustración) de forma que el alumno pueda reconocer que se trata de algo normal y tenga oportunidad de validar sus emociones sin reprimirlas para que en el futuro pueda controlarlas y así evitar que interfieran en su desempeño.

En segundo lugar, a través del *debriefing* se analizan los hechos de la práctica y su actuación en la misma, de esta forma se favorece en los estudiantes la reflexión de su actuación, más allá de lo que hizo bien o mal, es reconocer que puede mejorar, todo esto en un entorno

de evaluación formativa. Cuando la evaluación es sumativa se utiliza la ayuda de instrumentos, ya sea lista de cotejo o rúbricas, un ejemplo de esta evaluación es el Examen Clínico Objetivo y Estructurado (ECOE) donde los escenarios de simulación se encuentran dentro de algunas estaciones del examen.

La experiencia vivida en una simulación requiere una planeación para despertar la emocionalidad del estudiante (aprendizaje emocional) y, de esta manera, dejarlo plasmado como una experiencia novedosa y un aprendizaje perdurable para el estudiante.

Es de vital importancia, contar con docentes que sepan diseñar escenarios capaces de motivar a los estudiantes para activar su participación en el ambiente simulado, con la finalidad de lograr una identificación de su desempeño como parte de una realidad.

Para evaluar los resultados obtenidos de una simulación a través de escenarios se deben tomar en cuenta las siguientes fases del *debriefing*:

1) **Fase de emociones:** Expresión emocional de lo vivido en el escenario.
2) **Fase de hechos:** reconocimiento de lo que se vivió en el escenario.
3) **Fase de conceptualización:** Indagación y evocación de la simulación para conocer los esquemas mentales o puntos de vista de los participantes durante el escenario, los participantes exploran, analizan y reflexionan sus acciones y procesos del pensamiento, otra información para mejorar su rendimiento en situaciones reales.
4) **Fase de conclusión y cierre**: Esto con la finalidad de ayudar a los estudiantes sobre el desempeño que tuvieron, como ventanas de oportunidad para mejorar sus habilidades, seguidas del cierre de la sesión.

Cuando la actividad por simulación es por práctica deliberada, la evaluación empleada es a través de una retroalimentación la cual se considera una práctica donde el docente es una guía que promueve el desarrollo y potencial del alumno. Es definida como una habilidad del docente para compartir información específica con el alumno sobre su desempeño, para lograr que alcance su máximo potencial de aprendizaje de acuerdo con su etapa de formación y tiene el propósito de resaltar las diferencias entre el rendimiento real y el rendimiento esperado para generar un cambio de comportamiento.

LA SIMULACIÓN EN LA EVALUACIÓN A DISTANCIA

A partir de la pandemia por COVID-19 se incrementó la necesidad de continuar con el proceso de enseñanza-aprendizaje-evaluación con el apoyo de recursos en línea y estrategias que ayuden a los estudiantes a continuar con su formación y con el desarrollo de competencias que contribuirán a su aprendizaje y a su crecimiento profesional. Por esta razón, la simulación resulta una estrategia muy útil que facilita la aplicación de los conocimientos adquiridos en el espacio educativo bajo situaciones controladas y con la oportunidad de repetir la actividad varias veces.

En virtud de este contexto, la simulación permite la utilización de software y hardware para plantear al estudiante situaciones lo más semejante posible a la realidad e inclusive, realizar experimentos que promueven la comprensión de los contenidos establecidos en el programa académico. En este sentido la simulación virtual es una experiencia parcialmente inmersiva en un entorno digital (por ejemplo: computadora, tableta, teléfono, pantalla, etc.) que se utiliza para favorecer el aprendizaje, dirigida a resultados específicos; a través de Internet, tecnologías.

La simulación como herramienta de evaluación a distancia se puede implementar de dos formas:

a) **Sincrónica:** mediante plataformas de Zoom, GoogleMeet, entre otros, es la llamada "telesimulación" aquí se implementan escenarios virtuales, donde los estudiantes se enfrenta a contextos muy similares a la realidad, para resolver problemas, y ponen en práctica las habilidades de argumentación, análisis, razonamiento, reflexión, toma de decisiones y solución de problemas. Un ejemplo claro de esta variante es la "teleconsulta" o "telemedicina" que en los últimos años ha cobrado gran relevancia en el ámbito médico. La telesimulación emplea un elemento fundamental de la simulación en educación médica que es el paciente simulado. Este es una persona que se capacita en las artes escénicas para representar una enfermedad. La telesimulación también se puede operar con la ayuda de simuladores de alta fidelidad, donde las variantes fisiológicas (temperatura, pulso, respiración, etc.) son muy importantes para el escenario.
b) **Asincrónica:** esta se lleva a cabo de forma muy similar a un juego o a un videojuego en el que el estudiante tiene un papel central, ya que de él depende la toma de decisiones y, por lo tanto, el resultado que se obtenga a partir de sus elecciones, además de promover el trabajo colaborativo. Existe hoy en día, muchos y muy variados simuladores virtuales en todas las áreas de la ciencia. Desde los simuladores de vuelo, hasta los maniquíes de cuerpo completo que funciones fisiológicas antes mencionadas.
c) Aunado a las formas de implementación de la simulación virtual, existen diferentes tipos de simulación virtual como:

    1) **Telesimulación.** Sinónimos: Simulación por control remoto; simulación virtual sincrónica. Utiliza tecnología de información y comunicación (TIC), como las herramientas de videoconferencia, para proporcionar una simulación entre estudiantes e instructores ubicados de forma remota entre sí y de manera sincrónica.
    2) **Simulación basada en computadora.** Uso de un medio alternativo para la simulación donde los estudiantes completan tareas, toman decisiones y ven los resultados de las mismas.
    3) **Gamificación.** Integración de los principios de diseño de juegos para motivar a los estudiantes a participar en actividades educativas.

4) **Escenario de ramificación.** Un enfoque interactivo de simulación donde las decisiones de los estudiantes cambian la dirección de la historia, de esta forma se enfrentan a diferentes opciones.
5) **Camino lineal.** Un enfoque de la simulación virtual donde solo hay una opción correcta. Cuando un estudiante hace clic en una opción incorrecta se les presenta una justificación de por qué la opción era incorrecta y se le indica que vuelva a intentar la pregunta.
6) **Realidad Aumentada (AR).** Consiste en la integración de contenidos gráficos sobre una vista del mundo real de esta forma se combina la información digital conocida como capas de información digital como imagen, vídeo, audio, animaciones, texto, URLs y física en tiempo real a través de diferentes dispositivos tecnológicos como smartphones o tablets, para crear una nueva realidad enriquecida que permite la inmersión, la navegación e interacción.
7) **Realidad virtual (VR).** Es un entorno de escenas u objetos simulados de apariencia real que crean un ambiente inmersivo, interactivo, dinámico y adaptativo, en el que se emplean otros dispositivos como lentes, casco, auriculares donde se proyecta esa realidad.
8) **Vídeo 360.** Es un método de filmar en 360 grados para crear un cuadro completo del medio ambiente, se utiliza una cámara que puede filmar en todas las direcciones a la vez. Se pueden ver con un visor de realidad virtual, sin embargo, el entorno no es interactivo.
9) **Basada en casos o avatar.** Representación virtual de un ser humano a menudo capaz de mostrar respuestas físicas y expresiones faciales, lo que puede agregar alto impacto inmersivo.

¿Cómo diseño la simulación a distancia?

Para llevar a cabo actividades que involucren la simulación se requiere del apoyo de docentes expertos en el contenido que se va a evaluar y que colaboren en conjunto con expertos en producción de medios multimedia, según sea el caso.

Planeación

1) Identifique los objetivos de aprendizaje de acuerdo con su programa académico.
2) Defina el tipo de recursos que necesitará para llevar a cabo la simulación, además de contar con equipo de cómputo (software, maniquíes, personas, plataforma de videoconferencias, etcétera), considere si en su institución ya existen simuladores que se ajusten a lo que requiere evaluar y si es posible su utilización vía remota.
3) Establezca el tipo de evaluación ya sea diagnóstica, sumativa o formativa.
4) Elabore la descripción del problema, situación o actividad, así como el escenario en el que se va a realizar. Se requiere de expertos que apoyen la planeación y estructuración de la simulación.

Defina los criterios de evaluación, mismos que debe dar a conocer a los estudiantes antes de llevar a cabo la simulación.

Se sugiere establecer tiempos para la actividad e incluir un espacio para dar retroalimentación a los estudiantes.

> ***Recomendación:*** Considere los escenarios, situaciones o problemas más frecuentes a los que se puede enfrentar el estudiante y que se relacionan con el contenido que se abordará en la simulación.

¿Cómo instrumento la simulación a distancia?

Considere que en los espacios educativos la simulación debe estar elaborada de forma didáctica y cumplir con los objetivos de aprendizaje previamente establecidos.

Es importante garantizar que todos los estudiantes cuenten con equipo de cómputo apropiado para la actividad y estén familiarizados con la simulación que se va a realizar, en especial, si se trata de una evaluación sumativa.

Es necesario considerar que el nivel de complejidad de la simulación debe ser congruente con las características de los estudiantes y fácil de utilizar para que puedan adaptarse de forma óptima a la simulación.

Lo recomendable, si no se cuenta con experiencia en simulación, es comenzar con ejercicios sencillos que no requieran de software complejo para llevarlo a cabo a distancia.

También se sugiere que, al final de la simulación, los alumnos evalúen el proceso para que se identifiquen áreas de oportunidad y se puedan hacer modificaciones por parte de los alumnos y de los docentes a fin de mejorar la experiencia del aprendizaje.

Para comprender más acerca de las características de algunos modelos de simulación lo invitamos a ver el siguiente video que muestra un ejemplo aplicado a negocios: Modelos de simulación: https://www.youtube.com/watch?v=uBtHqqDvRDU

En caso de optar por la utilización de software más especializado, cabe mencionar que actualmente existen diversos recursos virtuales para llevar a cabo la simulación, a continuación, presentamos algunos ejemplos de tipos de simuladores utilizados de acuerdo con el propósito educativo:

Software de diálogo: Es un simulador que imita una conversación real con un cliente, colega o cualquier otra persona. Resulta muy útil para que los alumnos adquieran habilidades de comunicación. Entre los más conocidos se encuentran: iSpring Suite Max, Branch Track y Articulate 360.

Software para simulación de software: Representa un modelo del software que permite demostrar las características y funcionalidades. Resulta útil para aprender la utilización de algún software. Ejemplos de estos son Adobe Captive 2021 y Atomi ActivePresenter 8.

Software de simulación de comportamiento: Permiten imitar situaciones de la vida real para que los estudiantes practiquen sus habilidades que incluso pueden incluir simuladores de realidad virtual (VR, por sus siglas en inglés) o realidad aumentada (AR, por sus siglas en inglés). Ejemplos de estos son ITyStudio, Uptale, Vuforia Enterprise AR Suite.

Software para la simulación de sistemas:

Anylogic

Es un software gratuito de modelado de simulación con fines educativos dirigido a académicos, estudiantes y especialistas de la industria. Combina los métodos de eventos discretos, dinámica de sistemas y simulación basada en agentes.

HYSYS

Es un programa interactivo enfocado a la ingeniería de procesos y toda clase de problemas relacionados con procesos químicos que cuenta con una interfaz muy amigable para el usuario, además de permitir el empleo de operadores lógicos y herramientas que facilitan la simulación de diversos procesos.

ProModel

Es un programa de simulación de procesos industriales, permite simular cualquier tipo de proceso de manufactura, además de procesos logísticos, procesos de manejos de materiales y contiene simulaciones de talleres, grúas viajeras, bandas de transporte. entre otros.

Body Interact

Es un simulador clínico virtual, promueve el desarrollo de la competencia para establecer desde un diagnóstico clínico hasta el tratamiento, e integrar conceptos científicos básicos, protocolos médicos y de enfermería con la resolución de problemas clínicos en pacientes virtuales. (https://medical-simulator.com/simulacion-clinica-virtual/2457-body-interact.html)

Laerdal Medical LLEAP

Es un software de uso abierto, que permite la visualización de los signos vitales y los resultados de las pruebas de diagnóstico entre otros.

EJERCICIO INTERACTIVO:

En la siguiente dirección se puede tener acceso a algunos laboratorios virtuales o simuladores utilizados por universidades en todo el mundo y puede seleccionar por área o disciplina: Laboratorios virtuales/simuladores https://www.itba.edu.ar/intranet/clasesonline/laboratorios-virtuales-simuladores/. Le proponemos explorarlo y realizar alguna actividad de simulación.

¿CÓMO ANALIZO SUS RESULTADOS?

Al analizar los resultados se debe tomar en cuenta, además de los aprendizajes adquiridos, el progreso de cada estudiante respecto a las habilidades requeridas en la simulación.

En los ejercicios de simulación cuyo propósito de evaluación es formativo es importante dar retroalimentación a los alumnos para brindarles la oportunidad de rectificar los errores.

Los instrumentos específicos que se utilizaron para la autoevaluación, coevaluación, reinformación y/o reflexión entre los participantes contribuyen a lograr objetividad, equidad y precisión.

Al igual que con otras estrategias de evaluación, se pueden utilizar rúbricas o listas de cotejo para evaluar el desempeño de los estudiantes durante la evaluación.

EJEMPLOS

Ejemplo 1

- La simulación ha sido empleada con frecuencia en diversas áreas como una estrategia eficaz en la formación de los estudiantes. Debido a las ventajas que ofrece en la formación de los alumnos la han hecho idónea para distintos campos disciplinares, por ejemplo, en el área de Ciencias Sociales.
- El ejemplo que se presenta a continuación está dirigido a alumnos que pretenden obtener un certificado que avale que son capaces de pilotear drones.

Figura 1. Piloteo de un dron



1. Curso para aprender las habilidades prácticas necesarias para controlar y pilotear drones.

Objetivo: ser capaz de controlar de una aeronave (dron) y así obtener el certificado práctico de "piloto de drones".

Para lograr el control y las habilidades prácticas para ser piloto de drones es recomendable comenzar a practicar a través de los simuladores, mediante los cuales se pueden realizar los ejercicios como si los aspirantes estuvieran en el campo con su aeronave, pero de una forma más controlada y segura.

En la actualidad, existen muchas opciones de simuladores, pero todos cuentan con lo básico respecto a elementos y funcionalidades.

Evaluación de la simulación

A continuación, se presenta una rúbrica con la que se evaluará el desempeño del aspirante a piloto de drones.

Instrucciones para el evaluador: Responda el siguiente instrumento de evaluación de acuerdo con el desempeño observado.

Tabla 1. Rúbrica para evaluar el desempeño del aspirante a piloto de drones

Instrucciones: Elija una de las opciones de acuerdo con el desempeño del sustentante.

EVALUAR: El sustentante realizó o demostró:	VALORACIÓN		
Un despegue vertical seguido de un vuelo de 10 segundos en estacionario a la altura de los ojos del piloto y a una distancia de 5-10 metros del mismo.	Deficiente	Regular	Satisfactorio
Una traslación en vuelo rápido y nivelado en alejamiento en forma de S con cuatro cambios de rumbo a una altura de 20-30 metros.	Deficiente	Regular	Satisfactorio
Un vuelo lento de traslación nivelado en acercamiento, moviéndose hacia atrás (con la cola a la vista del piloto) y a 20/30 metros de altura.	Deficiente	Regular	Satisfactorio
Una traslación en vuelo lento y nivelado en alejamiento en forma de S con cuatro cambios de rumbo a una altura de 20-30 metros.	Deficiente	Regular	Satisfactorio
Un ascenso vertical a 10-20 m. de altura y un vuelo de traslación lateral a 30 m de distancia del piloto.	Deficiente	Regular	Satisfactorio
Ascenso a 50-60 m de altura y un viraje de 360º en descenso a la derecha (espiral), hasta 5 m de altura quedando frente al piloto.	Deficiente	Regular	Satisfactorio
Ascenso a 50-60 m de altura y un viraje de 360º en descenso a la izquierda (espiral), hasta 5 m de altura quedando frente al piloto.	Deficiente	Regular	Satisfactorio
Un circuito rectangular hacia la izquierda con aterrizaje vertical delante del piloto.	Deficiente	Regular	Satisfactorio
Un circuito a 100 m de altura y aterrizaje de emergencia a 50 m. del piloto.	Deficiente	Regular	Satisfactorio
Despegue y circuito rectangular con virajes a la izquierda con tramo final en acercamiento a 5 m de altura aterrizando en una zona definida a 10 m. del piloto.	Deficiente	Regular	Satisfactorio
La capacidad del multirrotor (pérdida de señal GPS, pérdida de enlace de mando y control), sistema RTH.	Deficiente	Regular	Satisfactorio
Un ascenso vertical a 100 metros de altura y descenso en vertical con aterrizaje.	Deficiente	Regular	Satisfactorio

Obtenga el puntaje total de la prueba de acuerdo con los siguientes valores:
Deficiente: 1 punto
Regular: 3 puntos
Satisfactorio: 5 puntos
Puntaje ≤ 30= No acreditado
Puntaje 31 a 50= Acreditado con conocimientos básicos
Puntaje 51 a 60= Acreditado con conocimientos avanzados

Ejemplo 2

2. Este ejemplo aborda la utilidad de la simulación en un curso de reanimación, ya que se evalúa el manejo de un paro cardiorrespiratorio.

El paro cardiorrespiratorio (PCR) representa una importante causa de muerte a nivel mundial, con una incidencia de 33 a 50 por cada 100,000 habitantes y con una sobrevida global que no supera el 8%. La intención de mejorar la respuesta y la sobrevida frente a un episodio de PCR es un tema de gran interés y depende de varios factores, entre ellos la disposición del espectador para realizar las maniobras de reanimación cardiopulmonar (RCP) y el nivel de capacitación del rescatista.

Desde la descripción del masaje cardiaco externo y la ventilación boca a boca en 1960, la difusión de estas técnicas ha ido progresando con el tiempo.

Así, en 1973, la American Heart Association (AHA) apoyó la formación de la población general, para posteriormente, en 1974, introducir el soporte vital cardiovascular avanzado (ACLS).

Se reconoce que la utilización de maniquíes puede servir para la integración de los conocimientos, habilidades y comportamientos en situaciones en las que se requiera reanimación cardiopulmonar.

Figura 2. Reanimación cardio-pulmonar en el adulto



A pesar de los importantes avances realizados en la prevención, el paro cardíaco continúa siendo un problema de salud pública significativo y una de las principales causas de muerte en muchos países del mundo. El paro cardíaco se produce tanto dentro como fuera del hospital. Por lo tanto, el adecuado manejo de estas situaciones tras la creación de esquemas de actuación estandarizada es fundamental en la formación del médico y la población en general.

Objetivo: evaluar en el alumno las habilidades de RCP para víctimas adultas, así como reconocer emergencias tales como un paro cardíaco súbito y socorrer a víctimas de las mismas.

Simulación

Desarrollo de la práctica:

Los alumnos deberán presentarse puntualmente a la hora reservada.

Material:

- modelo de tórax adulto
- ambú
- mascarillas
- guantes
- filtro boquilla
- DEA (desfibrilador)

Procedimiento:

1) Se dará un espacio de 30 minutos para recordar los pasos de RCP puede realizarse a través de la práctica deliberada (Identificación de la víctima, activación del sistema de emergencia, RCP de calidad, desfibrilación rápida, y cuidados postparto).
2) Posteriormente manera individual se reproducirá un escenario clínico habitual, donde se evaluarán todos estos pasos de forma individual y con apoyo.

Mediante la siguiente lista de cotejo se pretende evaluar el conjunto de habilidades necesarias para socorrer a una persona que presente paro cardiorrespiratorio en la vía pública.

Se plantea un escenario hipotético donde una persona (maniquí), al parecer del sexo femenino en la sexta década de la vida que va caminando, se lleva las manos al pecho y cae súbitamente al piso.

Realice las acciones necesarias ante este caso:

Tabla 2. Evaluación de actividades en la reanimación cardio pulmonar de un adulto

CONDUCTA O PROCEDIMIENTO	SI	NO
1. Verifica si la zona es segura.		
2. Comprueba estado de conciencia (EC) de manera verbal.		
3. Comprueba EC mediante estímulo táctil en los hombros del paciente.		
SI EL PACIENTE NO RESPONDE:		
4. Activa sistema de respuesta a emergencia (SRE) al "comprometer" a una persona en específico.		
5. Le especifica a una persona que llame al número 911.		
6. Le solicita una ambulancia con DEA.		
8. Descubre el tórax.		
9. Revisa respiración o movimiento de tórax y pulso central simultáneamente por 5-10 segundos, contando en voz alta.		
SI NO RESPIRA Y NO TIENE PULSO:		
10. Inicia compresiones efectivas con posición correcta de manos.		
11. Realiza ciclos de RCP con ritmo de 30 compresiones por 2 ventilaciones.		
12. El evaluador cronometra cada ciclo y comprueba que es <18´ (significa que da mínimo 100 y máximo 120 compresiones por minuto).		
13. Las compresiones tienen mínimo 5 cm de profundidad, máximo 6 cm.		
14. Permite elevación torácica entre compresiones (re-expansión torácica).		
15. Realiza maniobra frente-mentón.		
16. Ocluye las narinas.		
17. Administra dos ventilaciones.		
18. Verifica elevación del tórax.		
19. Minimiza interrupciones por menos de 10 segundos.		
20. Realiza un radián completo (5 ciclos 30x2)		
21. Revisa pulso central por 5 a 10 segundos.		
SI NO TIENE PULSO NI RESPIRA:		
22. Realiza un radián completo efectivo.		
23. Revisa el pulso central.		
Observaciones:		

RECURSOS EN LÍNEA

Tabla 3. Recursos en línea

Recurso	Descripción
Mundo 4.0 (Simuladores educativos) https://youtu.be/_1KECC4oufM	En este video se describen las características y la aplicación de los simuladores educativos.
Clases a distancia usando Simulaciones Interactivas PhET https://youtu.be/Hs9qYSTBA_I	En este video se aborda el tema de las clases a distancia (relacionado con la situación de la pandemia por el COVID-19 usando simulaciones interactivas PhET por parte de una profesora que es integrante de la Universidad de Colorado Boulder, Estados Unidos.
El uso de la Simulación en Aprendizaje Virtual https://recursos.educoas.org/publicaciones/el-uso-de-la-simulaci-n-en-aprendizaje-virtual	En este documento se describen las ventajas de emplear la simulación a distancia para el desarrollo de la toma de decisiones y otras habilidades que se requieren para la formación profesional de los estudiantes.
Simuladores virtuales: recursos de apoyo en el proceso de aprendizaje https://igniteonline.la/simuladores-virtuales-recursos-de-apoyo-en-el-proceso-de-aprendizaje/	Página electrónica que detalla la utilidad de la simulación en el proceso de aprendizaje. Además, contiene in link para acceder al webinar "Práctica de laboratorio virtual: separación de mezclas (repetición)".

CONCLUSIONES Y REFLEXIONES FINALES

- La utilización de la simulación como técnica para el aprendizaje y la evaluación continúa extendiendo su campo de acción en áreas diversas en las que se requiere que los alumnos apliquen los conocimientos teóricos adquiridos, a fin de desarrollar destrezas y habilidades que utilizarán en su práctica profesional.
- En este tipo de instrumento de evaluación es necesario que el docente considere a la retroalimentación como un elemento indispensable para la formación de los alumnos.
- La simulación con escenarios bien diseñados ya sea que se realice de forma presencial o a distancia, contribuirá de manera eficaz a que el alumno adquiera y/o mejore sus capacidades en toma de decisiones, procedimientos, además de valores como el compromiso y la responsabilidad.

Se sugiere que la elección de simuladores sea lo más parecida a escenarios reales a los que se enfrentarán los estudiantes en el futuro, ya que esto es indispensable para lograr el nivel de competencia esperado en el ámbito profesional.

Es recomendable que la simulación esté integrada al currículo para mantener una coherencia con los objetivos y resultados de aprendizaje de los programas académicos.

La simulación es una técnica de gran impacto en la educación, cuando se utiliza con otras metodologías activas como el Aprendizaje Basado en Proyectos, método de caso, entre otros.

REFERENCIAS

Arriaza N. y Rocco C. (2012). Educación y simulación en reanimación cardiopulmonar. *Revista Chilena de Anestesia, 41*, 53-56.

Barrientos M. (2017). Práctica # 5.1: Reanimación cardio-pulmonar en el adulto. Unidad Médica de Simulación Clínica "Dr. José Jorge Talamas Márquez». Durango, México: Facultad de Medicina y Nutrición de la Universidad Juárez del Estado de Durango. Recuperado de: https://docplayer.es/33211305-Practica-5-1-reanimacion-cardio-pulmonar-en-el-adulto.html

Cabero-Almenara J & Costas J. (2016). La utilización de simuladores para la formación de los alumnos. *Prisma Social, 17*, 343-372. [fecha de Consulta 8 de octubre de 2021]. Recuperado de https://www.redalyc.org/articulo.oa?id=353749552015

"Debriefing" y evaluación de la simulación. (2012). Asociación de simuloeducadores de Puerto Rico. Recuperado de https://www.academia.edu/16746044/Debriefing-y-Evaluacion-en-la-Simulacion

Dolan H, Amidon BJ & Gephart SM. (2021). Evidentiary and theoretical foundations for virtual simulation in nursing education. *Journal of Professional Nursing, 37*, 810–815. Recuperado de https://doi.org/10.1016/j.profnurs.2021.06.001

Foronda CL. (2021). What Is Virtual Simulation? *Clinical Simulation in Nursing, 52*. 8. Recuperado de https://doi.org/10.1016/j.ecns.2020.12.004

Hernández, L. (2021). *Evaluación para el aprendizaje: enfoque cualitativo. Simulación* [MOOC] https://www.coursera.org/lecture/evaluacion-cualitativa/simulacion-mSkg8

Gaba DM & De Anda A. (1988). A comprehensive anesthesia simulation environment: re-creating the operating room for research and training. *Anesthesiology, 69*, 387-94.

Gaba D. (2004). The future vision of simulation in health care. Qual Saf Health Care, 13, i2–10.

Harder N. (2010). Use of Simulation in Teaching and Learning in Health Sciences: A Systematic Review. *Journal of Nursing Education, 49*, 23-28.

Kuszajewski ML, Vaughn J, Bowers MT, et al. (2021). Embracing Disruption: Measuring Effectiveness of Virtual Simulations in Advanced Practice Nurse Curriculum. *Clinical Simulation in Nursing, 57*, 41–47. Recuperado de https://doi.org/10.1016/j.ecns.2021.04.017

McDermott D, McEwing E, Lopez J, Osso M, et al. (2021). Pandemic Crisis: Simulation Contingency Plans. *Teaching and Learning in Nursing, 16*, 393–395. Recuperado de https://doi.org/10.1016/j.teln.2021.04.007

Pottle J. (2019). Virtual reality and the transformation of medical education. *Future Healthcare Journal*, *6,* 181–185. Recuperado de https://doi.org/10.7861/fhj.2019-0036

Recomendaciones de práctica para la obtención del certificado práctico de piloto de drones. (2017). Recuperado de https://www.youtube.com/watch?v=HyjtfU8Yhqk

Salas-Perea R. y Ardanza-Zulueta P. (1995). La simulación como método de enseñanza y aprendizaje. *Revista Cubana de Educación Médica Superior, 9*(1-2). Recuperado de http://www.bvs.sld.cu/revistas/ems/vol9_1_95/ems03195.htm

Secretaría de Educación Pública. (2017). Orientaciones para la prevención, detección y actuación en casos de abuso sexual infantil, acoso escolar y maltrato en las escuelas de educación básica. Documento base para elaboración de protocolos en las entidades federativas. Recuperado de http://www.gob.mx/cms/uploads/docs/Orientaciones_211216.pdf

Verdejo P., Encinas M. y Trigos L. (2003). Estrategias para la evaluación de aprendizajes complejos y competencias. Recuperado de http://www.innovacesal.org/innova_public_docs01_innova/ic_publicaciones_2012/pubs_ic/pub_03_doc03.pdf

Páginas WEB consultadas:

Anylogic. Software de simulación gratuito para educación. [Consultado el 28 de septiembre de 2021]. Recuperado de https://www.anylogic.com/s/software-de-simulacion-gratuito-para-educacion/

Softwares utilizados para la simulación de sistemas. [Consultado el 29 de septiembre de 2021]. Recuperado de http://softwaresdesimulacion.blogspot.com/2014/02/softwares-de-simulacion.html

Software-educativos. Los 8 mejores softwares de simulación para crear experiencias prácticas. (Internet) [Consultado el 28 de septiembre de 2021]. Recuperado de https://software-educativo.com/software-educativo-simuladores/

Capítulo 26

ENSAYO

María Azucena Montoya Magno, Ana Laura Pérez Díaz

“El leer hace completo al hombre, el hablar lo hace expeditivo,
el escribir lo hace exacto.”
Sir Francis Bacon

INTRODUCCIÓN

Si bien es cierto que, en las distintas etapas escolares, los estudiantes se enfrentan a la elaboración y construcción de distintos textos, es la universidad quien le ofrece una formación académica con situaciones comunicativas y géneros más especializados, lo cual forma parte de la alfabetización académica (Carlino, 2005, 2008, 2013; González e Ibáñez, 2017, citados en Osorio, et al., 2019)

En cuanto a la producción de textos académicos solicitados por los profesores, el ensayo es un recurso de evaluación muy utilizado porque permite obtener evidencia respecto al desarrollo de habilidades de pensamiento complejo como son: el desarrollo del pensamiento crítico, el pensamiento lógico y argumentativo, la capacidad de análisis, de síntesis y de investigación en los alumnos.

En los distintos espacios educativos, los ensayos se elaboran a partir de lecturas estudiadas en las clases o recomendadas en las unidades curriculares, y los profesores deben estar conscientes que la elaboración de un ensayo será una actividad que favorece otra competencia comunicativa, que es la escritura.

La escritura debe ser –y es una opinión compartida por la mayoría de los profesionales de la educación– una de las principales competencias que toda persona debe desarrollar, siendo esencial en el aprendizaje del resto de materias y entronca con el resto de las competencias formando un todo indivisible (Barrera, 2019).

Al ser la escritura parte de una competencia comunicativa clave para el desarrollo personal y profesional de los estudiantes, los profesores deben conocer qué aspectos evaluar de dicha competencia (Ruíz-Terroba, et al., 2017).

La evaluación del y para el aprendizaje es el proceso sistemático que permitirá obtener información respecto a lo que acontece en el proceso educativo, y uno de sus propósitos es

retroalimentar a los docentes y a los alumnos respecto a su desempeño con el objetivo de guiar y, de ser necesario, modificar su planeación de la enseñanza y el aprendizaje para que éstos sean significativos. Al evaluar un ensayo, el profesor debe enfocarse fundamentalmente tanto en la claridad de las ideas, como en la manera en que éstas se organizan, exponen y argumentan, con el propósito de verificar que se está desarrollando la competencia comunicativa y que se están cumpliendo los resultados de aprendizaje esperados.

El propósito de este capítulo es poner al alcance y guiar al profesor en la instrumentación de los ensayos en sus prácticas de la evaluación del y para el aprendizaje.

Además, es un recurso de autoestudio, reflexión y consulta que puede leerse o no en orden.

Este capítulo está organizado en seis secciones: en la primera se explica qué es, en la segunda se detallan los pasos para diseñarlo, en la tercera se recomienda cómo aplicarlo, además de que en esta sección se presenta información del desarrollo de un ensayo en un marco de la evaluación a distancia, para ello, se especifica cómo diseñarlo y aplicarlo en dicha modalidad. En la cuarta sección se puntualiza cómo analizar sus resultados. La quinta sección presenta ejemplos y recursos en línea de apoyo y, por último, en la sexta se presentan conclusiones y reflexiones finales.

¿QUÉ ES?

- El ensayo es un texto o composición original por parte del alumno, en la cual se desarrolla un tema a partir del conocimiento que tiene sobre este y su reflexión crítica, sustentando sus ideas y opiniones en evidencias teóricas y empíricas.
- En el contexto educativo destacan el ensayo de respuesta extensa y el ensayo de respuesta restringida.
- En el caso de los ensayos de respuesta extensa, que son los que abordaremos en este capítulo, de acuerdo con el propósito de los aprendizajes esperados que se requiera evaluar, los ensayos pueden ser de tres tipos: expositivo, argumentativo y analítico.
- El ensayo expositivo se puede utilizar para verificar en qué grado se han desarrollado las habilidades de investigación en los alumnos y si son capaces de emitir una opinión respecto a un tema.
- El ensayo argumentativo es recomendable cuando se requiere evaluar la articulación lógica y coherente de ideas sobre un tema, así como la capacidad de argumentación en los alumnos.
- El ensayo analítico, como su nombre lo indica, permite evaluar el desarrollo de la capacidad de análisis e interpretación en los alumnos.
- En el ámbito educativo, el ensayo, como un recurso de evaluación del y para el aprendizaje, es útil en la evaluación formativa o sumativa, porque permite valorar el grado en que los alumnos han desarrollado procesos cognitivos complejos, como la capacidad de organización, razonamiento, análisis, síntesis y argumentación, así como sus habilidades de investigación, de escritura y de comunicación de ideas de manera lógica y coherente. Además, este recurso favorece principalmente el desarrollo de la competencia

comunicativa de escritura, que es una clave para el desarrollo personal y profesional de los estudiantes.

- Generalmente se solicita a los alumnos que realicen el ensayo de manera individual, como parte de la evaluación formativa o sumativa, ya que esto permite obtener información sobre los aprendizajes logrados por el alumno en una unidad del programa o al finalizar el curso.
- Por su carácter reflexivo, el ensayo es un recurso utilizado más frecuentemente para evaluar conocimientos y habilidades en las áreas de las humanidades y las artes, así como en las ciencias sociales.
- Una de las ventajas del ensayo es que alienta a los estudiantes a la reflexión y al desarrollo del pensamiento crítico.
- Una de las desventajas del ensayo es que el contenido que se evalúa puede ser limitado, y, además, su calificación implica una gran cantidad de tiempo por parte de los profesores; asimismo es subjetiva, pues depende del punto de vista de cada profesor la puntuación asignada al ensayo.

Figura 1

¿CÓMO LO DISEÑO?

El ensayo como un recurso de evaluación del y para el aprendizaje, independientemente de que se solicite en modalidad presencial, a distancia o híbrida, requiere de una planeación sistemática y formativa, que se encuentre alineada a los objetivos de aprendizaje que se desea logre o domine el alumno.

A continuación, se mencionan los pasos a seguir para la planeación y diseño de un ensayo.

Planeación

1) **Seleccione los aprendizajes esperados que se evaluarán mediante el ensayo.**
   Analice el programa de la asignatura o unidad y seleccione los aprendizajes que requiere evaluar: ¿El ensayo es el mejor instrumento para valorar los aprendizajes esperados? ¿Es viable de acuerdo con las características del grupo y del tiempo con el que se cuenta? ¿Es viable de acuerdo con el tiempo disponible para realizar la revisión y evaluación?
2) **Determine el tipo de evaluación.**
   El ensayo puede utilizarse en la evaluación formativa, ya que permite retroalimentar a los alumnos respecto a los aspectos que se tienen que mejorar durante la elaboración del ensayo para que se logren los aprendizajes esperados en una unidad. Asimismo, también se usa en la evaluación sumativa para valorar el grado en que los alumnos han desarrollado las capacidades y habilidades del pensamiento complejo como el pensamiento crítico, la capacidad de análisis, síntesis, entre otras, al finalizar un curso.
3) **Especifique quiénes serán los participantes en la evaluación.**
   El profesor es el principal actor de la evaluación del ensayo (heteroevaluación), pues es quien establece los criterios de valoración del ensayo. Sin embargo, la evaluación también puede ser realizada por pares (coevaluación), quienes podrán retroalimentar al alumno en el proceso de elaboración siempre y cuando conozcan previamente los elementos que se valorarán.
4) **Especifique los criterios de evaluación del ensayo de acuerdo con los aprendizajes esperados.**
   Determine los criterios o elementos que valorará en el ensayo, por ejemplo: la claridad del tema, el planteamiento de la tesis que sustentará el desarrollo del ensayo, la articulación adecuada entre las partes del ensayo, la originalidad en la presentación de ideas, la capacidad de argumentación, síntesis, análisis, las ideas se presentan de manera coherente y lógica, las fuentes de información utilizadas, entre otros. Comparta dicha información con los alumnos.
5) **Defina las características del ensayo.**
   Especifique la extensión del ensayo y el tiempo con el que se contará para su elaboración y entrega. En caso de que el ensayo se solicite en la modalidad a distancia, es necesario que el profesor especifique la plataforma o el medio por el cual será entregado al profesor y donde posteriormente el alumno recibirá la retroalimentación.
6) **Solicite el apoyo de un colega para que revise los criterios que se valorarán en el ensayo.**
   ¿Los criterios a valorar corresponden con los resultados de aprendizaje que se requieren evaluar en los alumnos? ¿Son claros y precisos?
7) **Elija los instrumentos que se utilizarán para la evaluación del ensayo.**
   Para que la evaluación del ensayo se lleve a cabo de manera sistemática, se recomienda el uso de instrumentos tales como la rúbrica o la lista de cotejo, pues ambos permiten dar cuenta de si se han logrado los aprendizajes esperados y si se han cumplido con los criterios establecidos.

Figura 2

Planeación

Seleccione los aprendizajes que se evaluarán mediante el ensayo

Determine el tipo de evaluación y los participantes

Defina las características del ensayo

Especifique los criterios de evaluación del ensayo

Elija los instrumentos que se utilizarán para la evaluación del ensayo

Paso 1 | Paso 2 | Paso 3 | Paso 4 | Paso 5

Diseño

1) **Defina el tema del ensayo.**
Se recomienda que sea corto, claro, preciso, sintético y atractivo.
2) **Incluya una sección de datos de identificación del alumno.**
Nombre de la asignatura, datos del profesor, datos del alumno y fecha.
3) **Redacte las instrucciones para la elaboración del ensayo.**
Explique con claridad y precisión las características del ensayo como la extensión, formato, tiempo para su elaboración, entrega y criterios a evaluar, entre otros.
4) **Elabore la lista de cotejo o rúbrica que utilizará para la valoración del ensayo.**
En la rúbrica o lista de cotejo se expresarán los elementos o criterios a valorar en el ensayo. Por ejemplo: claridad en la presentación de ideas, organización de ideas, capacidad de análisis y de argumentación, fuentes consultadas, entre otras, que estarán en función de los aprendizajes que se espera demuestren los alumnos.
5) **Proporcione a sus colegas la rúbrica o lista de cotejo con los elementos a valorar en el ensayo para su revisión.**
Solicite a sus colegas apoyo para revisar los criterios o lineamientos que se valorarán en el ensayo mediante la rúbrica o lista de cotejo, con el propósito de contar con sugerencias u observaciones que permitan mejorar el instrumento de evaluación.
6) **Realice los ajustes necesarios a la rúbrica o lista de cotejo que se utilizará para valorar los elementos en el ensayo.**
Con base en las observaciones y sugerencias de sus colegas, se recomienda realizar los ajustes que considere pertinentes para que el instrumento cumpla el propósito de la evaluación.

Revisión

Se sugiere que el profesor realice revisiones periódicas sobre el avance de la elaboración del ensayo con la finalidad de retroalimentar a los alumnos y verificar que se están apegando tanto a los resultados de aprendizaje esperados como a los criterios de evaluación establecidos.

¿CÓMO LO APLICO?

- El ensayo como recurso de evaluación puede utilizarse cuando se requiere que los alumnos seleccionen, organicen o integren materiales o ideas respecto a un tema específico.
- Cuando se requiere evaluar la comprensión de conceptos y sus habilidades de escritura.
- Se puede aplicar cuando se requiere evaluar la capacidad de los alumnos para construir argumentos sólidos y defendibles mediante razones lógicas frente a un grupo.
- Puede apoyar a la evaluación formativa para identificar las fortalezas y las áreas de oportunidad de los alumnos respecto a los aprendizajes esperados, con el propósito de implementar acciones educativas que permitan el logro de estos.
- Puede realizarse de manera individual o grupal, dentro o fuera del espacio educativo.
- Puede aplicarse al finalizar una unidad o un curso.
- Es necesario que, en todo momento, se tengan presentes los criterios que se considerarán en la evaluación del ensayo, con la finalidad de evitar juicios subjetivos durante el proceso de evaluación.

El ensayo en evaluación a distancia

En la actualidad, ante la situación de pandemia de COVID-19, los modelos educativos han tenido que adaptar y ajustar sus estrategias pedagógicas diseñadas para una enseñanza presencial, a una modalidad de educación en línea o híbrida. Este cambio resultó ser significativo ya que se ha privilegiado el desarrollo y aplicación de estrategias, técnicas y recursos que favorezcan el desarrollo de competencias que serán utilizadas por los estudiantes a lo largo de su vida en su ámbito personal, académico, social y profesional, entre ellas, la expresión escrita. (Ruiz-Terroba, et al, 2017).

- El ensayo como estrategia de aprendizaje puede utilizarse para fomentar el desarrollo de la competencia comunicativa de expresión escrita sin importar la modalidad en la que se solicite.
- El ensayo como recurso de evaluación puede utilizarse cuando se requiere que los alumnos seleccionen, organicen o integren materiales o ideas respecto a un tema específico. También para evaluar la comprensión de conceptos.
- Se puede aplicar cuando se requiere evaluar la capacidad de los alumnos para construir argumentos sólidos y defendibles mediante razones lógicas frente a un grupo.
- Puede apoyar a la evaluación formativa para identificar las fortalezas y las áreas de oportunidad de los alumnos respecto a los aprendizajes esperados, con el propósito de implementar acciones educativas que permitan el logro de estos.

- Puede aplicarse al finalizar una unidad o un curso.
- En la actualidad se puede hacer uso de herramientas tecnológicas para establecer la comunicación en el espacio educativo entre el profesor y los alumnos como Zoom, Google Meet, Cisco Webex Meetings, Blackboard, Microsoft Teams, entre otros, y en estos espacios puede solicitarse de manera síncrona o asíncrona la elaboración de un ensayo como actividad para evaluar ciertos resultados de aprendizaje.
- Recuerde que la elaboración de un ensayo puede solicitarse en una modalidad presencial, a distancia o híbrida.
- Es necesario que, en todo momento, se tengan presentes los criterios que se considerarán en la evaluación del ensayo, con la finalidad de evitar juicios subjetivos durante el proceso de evaluación.

¿Cómo lo diseño a distancia?

El diseño de un ensayo debe seguir un proceso de planificación cuidadoso, orientado a propósitos específicos, independientemente de la modalidad educativa en la que se solicite (presencial, a distancia o híbrida).

Si el ensayo se solicita de manera presencial en un espacio educativo, o se solicita a través de una videoconferencia síncrona, o de manera asíncrona, se sugiere establecer en todo momento los criterios de evaluación para darle mayor objetividad a la estrategia solicitada.

Diseño de la rúbrica de calificación para el ensayo en educación a distancia

Si el ensayo es visto como técnica de evaluación y se utiliza una rúbrica para evaluarlo, diséñela para orientar a los alumnos en el nivel de logro que se desea alcanzar, es decir, incluya los criterios y descripciones que serán evaluados y compártalos previamente. Recuerde que lo importante es definir los atributos y que estos sean congruentes con el aprendizaje esperado. Para conocer con mayor detalle el proceso de elaboración de la rúbrica consulte el capítulo 15 de este libro.

Consulte el video "Ensayo" Youtube https://youtu.be/CMNtfZ9ub3Y

¿Cómo lo instrumento?

- El ensayo puede solicitarse de manera presencial, a distancia o híbrida.
- Es importante que desde la planeación considere qué aprendizajes quiere evaluar a través del ensayo, para así orientar su elaboración.
- Es importante establecer las fechas establecidas para la elaboración, entrega, evaluación y retroalimentación del ensayo.
- La elaboración de un ensayo puede programarse para evaluar un tema al inicio, durante o al final de un curso y tener propósitos diferentes.
- Se recomienda tener listos los formatos con los criterios de evaluación del ensayo ya sea de manera impresa o digital. En cualquier caso, integre un apartado final para incluir

observaciones generales del ensayo que le serán útiles al momento de compartir la retroalimentación a los alumnos.

- Al concluir la actividad considere un espacio para retroalimentar a los alumnos. En la retroalimentación se sugiere mencionar los aspectos que el alumno domina y aquellos en los que deben mejorar, de acuerdo con los criterios de evaluación establecidos.

¿CÓMO ANALIZO SUS RESULTADOS?

De manera teórica, se puede evaluar la producción escrita en dos dimensiones: global y local. El aspecto global se refiere a la organización y estructura del texto y el aspecto local, hace referencia a elementos tales como la conexión entre ideas, ortografía y gramática. Estas dimensiones contribuyen a la riqueza argumentativa del texto (van Eemeren, 2001 citado en Muñoz, 2015). Errázuriz (2019) menciona algunos aspectos que se pueden evaluar en la elaboración de un ensayo: contenido, organización de ideas, cohesión, tesis, argumentos, contraargumentos, bibliografía utilizada, redacción, gramática, ortografía y vocabulario. Dependiendo del objetivo de aprendizaje y de la clase de razonamiento o resolución de problemas que el profesor quiera que demuestren los alumnos en la producción del ensayo, se deberán alinear los criterios que se evaluarán. El profesor deberá guiar al alumno durante el proceso de elaboración del ensayo y ser claro en los aspectos a evaluar; por lo tanto:

- El ensayo, como recurso de evaluación del y para el aprendizaje, debe contar con criterios o elementos previamente establecidos que permitan llevar a cabo tanto una valoración cualitativa respecto a la capacidad de organización, análisis, síntesis y argumentación de los alumnos, así como también de sus habilidades de escritura, comunicación de ideas de manera lógica y coherente, y sus habilidades para la investigación; esto es lo que se espera demuestren los alumnos.
- Para llevar a cabo el análisis de los resultados obtenidos por los alumnos en un ensayo, lo primero que debe considerarse es identificar por quién fue realizada la evaluación, si fue por el profesor –heteroevaluación– o los alumnos –coevaluación– y conocer el instrumento que se utilizó para la valoración del ensayo; en este caso, se puede utilizar una rúbrica o una lista de cotejo.
- En el caso de que para la evaluación del ensayo se haya utilizado una rúbrica holística, que incluya una escala de valoración de cuatro o cinco niveles de desempeño, se podrá inferir el desempeño del alumno, dependiendo del cumplimiento de las descripciones de cada uno de los elementos o criterios a evaluar, como: excelente, satisfactorio, suficiente o deficiente.
- Cuando se haya utilizado una rúbrica analítica, en la cual se presenten ciertos criterios a evaluar en el ensayo a los que se les ha asignado previamente un puntaje, se puede sumar el puntaje que se haya otorgado a cada criterio o aspecto para obtener un puntaje global y de esa manera se podrá realizar una inferencia sobre el desempeño del alumno.

- Si se ha empleado una lista de cotejo, este recurso permitirá verificar la presencia o ausencia de los criterios que se valoraron en un ensayo y, dependiendo del número de criterios que se cumplan, podrá obtener información respecto al desempeño del alumno y verificar si se han logrado los aprendizajes esperados.
- A partir del análisis de los resultados obtenidos mediante alguno de estos instrumentos de evaluación, el profesor podrá determinar las fortalezas y los aspectos en los que se requiere trabajar con los alumnos para lograr los aprendizajes que se esperan de ellos.
- Es importante destacar que la evaluación de un ensayo implica un mayor tiempo para la calificación que otros recursos de evaluación debido a que no se cuenta con una respuesta única, por lo que se sugiere contemplar un lapso más amplio para revisarlo, calificarlo y analizar sus resultados.

EJEMPLOS Y RECURSOS EN LÍNEA

Ejemplos

Tabla 1. Rúbrica para evaluar un ensayo

La siguiente rúbrica tiene como propósito realizar una valoración sistemática de un ensayo.

Instrucciones: A continuación, se presentan una serie de aspectos (introducción, tema, tesis, capacidad de organización, análisis, síntesis y argumentación; habilidades de escritura, ideas estructuradas de manera lógica y coherente, habilidades de investigación y fuentes de información consultadas) que deberá contener el ensayo para su evaluación. Por favor indique en la sexta columna el puntaje que asigna a cada uno de estos aspectos de acuerdo con la información presentada por el alumno.

Aspectos por evaluar en el ensayo	Excelente (4 puntos)	Satisfactorio (3 puntos)	Suficiente (2 puntos)	Deficiente (1 punto)	Puntaje asignado
Introducción	La información de la introducción despierta el interés del lector, explica de manera clara el objetivo del ensayo y menciona los subtemas o capítulos que se desarrollarán en el ensayo.	La información de la introducción despierta el interés del lector, explica de manera clara el objetivo del ensayo, pero no menciona los subtemas o capítulos que se desarrollarán en el ensayo.	La información de la introducción despierta el interés del lector, sin embargo, explica de manera ambigua el objetivo del ensayo y no menciona los subtemas o capítulos que se desarrollarán en el ensayo.	La información de la introducción es confusa, no despierta el interés del lector, explica de manera ambigua el objetivo del ensayo y no menciona los subtemas o capítulos que se desarrollarán en el ensayo.	
Tema	El tema y el objetivo del ensayo se expresan de manera clara y precisa, asimismo son congruentes con el contenido del ensayo. Se explica de manera clara y precisa el contenido que se desarrollará en el ensayo.	El tema y el objetivo del ensayo se expresan de manera clara y precisa, sin embargo, se explica ambiguamente el contenido que se desarrollará en el ensayo.	El tema del ensayo se expresa de manera clara y precisa; sin embargo, el objetivo del ensayo es confuso y no se explica claramente el contenido que se desarrollará en el ensayo.	El tema y el objetivo del ensayo no se expresan de manera clara y precisa y tampoco se explica claramente el contenido que se desarrollará en el ensayo.	

Tesis	La tesis del ensayo se expresa de manera clara y precisa en una sola idea afirmativa.	La tesis del ensayo se expresa de manera poco clara, con algunas imprecisiones, debido a que hay varias ideas contenidas en ella.	La tesis del ensayo no se expresa de manera clara, hay varias imprecisiones por la cantidad de ideas contenidas en ella.	No se expresa la tesis del ensayo, y solo se presentan varias ideas confusas que no permiten identificar la tesis sobre la que se sustentará el ensayo.	
Capacidad de análisis	Existe una articulación adecuada entre las partes que integran el ensayo y las conexiones entre ellas son precisas y claras.	Existe una articulación adecuada en la mayoría de las partes que integran el ensayo y las conexiones entre ellas son precisas y claras.	Existe una articulación en algunas partes que integran el ensayo, las conexiones entre ellas son poco precisas y claras.	No existe una articulación adecuada entre las partes que integran el ensayo y las conexiones entre ellas no son precisas, ni claras.	
Capacidad de síntesis	Las ideas que se presentan son originales, relevantes y se encuentran estructuradas de manera coherente y lógica.	Solo algunas de las ideas que se presentan son originales, relevantes, sin embargo se encuentran estructuradas de manera coherente y lógica.	Solo algunas de las ideas que se presentan son originales, se encuentran estructuradas de manera poco coherente y lógica.	La mayoría de las ideas que se presenta son irrelevantes y no son originales. Además, no están estructuradas de manera coherente y lógica.	
Capacidad de argumentación	Los argumentos presentados sustentan la tesis del ensayo. Las ideas presentadas son propias y están sustentadas en fuentes de información confiables que tienen relación con el tema.	Los argumentos presentados sustentan la tesis del ensayo, sin embargo, la mayor parte de las ideas presentadas no son propias, pero si están sustentadas en fuentes de información confiables.	Los argumentos presentados no sustentan la tesis del ensayo, la mayoría de las ideas no son propias, pero si están sustentadas en fuentes de información confiables.	Los argumentos presentados no sustentan la tesis del ensayo. Las ideas presentadas no son propias y en algunos casos están sustentadas en fuentes de información poco confiables.	

Habilidades de comunicación escrita	Las ideas en el ensayo se expresan de manera clara, coherente y lógica, y están construidas y escritas correctamente. Las transiciones entre los párrafos son adecuadas. El vocabulario empleado en el desarrollo del ensayo es adecuado para la audiencia a la que está dirigido.	La mayoría de las ideas en el ensayo se expresa de manera clara, coherente y lógica, asimismo están escritas correctamente. La mayoría de las transiciones entre los párrafos es adecuada. El vocabulario empleado en el desarrollo del ensayo es conveniente para la audiencia a la que está dirigido.	Algunas ideas en el ensayo se expresan de manera clara, coherente y lógica, asimismo están escritas correctamente. Algunas de las transiciones entre los párrafos son adecuadas. El vocabulario empleado en el desarrollo del ensayo es poco conveniente para la audiencia a la que está dirigido.	Las ideas en el ensayo no se expresan de manera clara, coherente y lógica, ni están escritas correctamente y son confusas. Las transiciones entre los párrafos no son adecuadas. El vocabulario empleado en el desarrollo del ensayo no es conveniente para la audiencia a la que está dirigido.	
Fuentes de información	Las fuentes de información consultadas son confiables, diversas, relevantes y actuales, asimismo se encuentran relacionadas con el tema.	Las fuentes de información consultadas son confiables, aunque no muy diversas, algunas relevantes y actuales, además se encuentran relacionadas con el tema.	Las fuentes de información consultadas son poco confiables y diversas. Algunas son relevantes y actuales, se encuentran relacionadas medianamente con el tema.	Las fuentes de información consultadas no son confiables, ni actuales y son muy pocas. La mayoría de ellas no se encuentran relacionadas con el tema.	

Fuente: Elaboración propia

Tabla 2. Lista de cotejo para la evaluación de ensayos

Esta lista tiene como propósito realizar una valoración sistemática de un ensayo.

Instrucciones: A continuación, se presentan una serie de elementos que deberá de contener el ensayo para su evaluación. Por favor señale con una X si el ensayo cumple o no con el elemento mencionado.

ELEMENTO A VALORAR	Cumple	
	Sí	No
Introducción		
La información de la introducción despierta el interés del lector y atrae su atención.		
Expresa de manera clara el planteamiento del tema del ensayo.		
Explica de manera clara el objetivo del ensayo.		
Menciona los subtemas o capítulos que se desarrollarán en el ensayo.		
Desarrollo		
Tema		
El tema está formulado de manera clara y precisa.		
El tema es congruente con el contenido del ensayo.		
El objetivo del ensayo está formulado de manera clara y precisa.		
Explica de manera clara y precisa el contenido del ensayo.		
Tesis		
La tesis del ensayo está expresada de manera precisa y en una sola idea.		
Se identifica de manera clara la tesis sobre la que se sustentará el ensayo.		
Capacidad de análisis		
Existe una articulación adecuada entre las partes que integran el ensayo.		
Las conexiones entre las partes que integran el ensayo son precisas, adecuadas y claras.		
Capacidad de síntesis		
Las ideas presentadas en el ensayo son propias y originales.		
Las ideas presentadas son relevantes y se expresan de manera coherente.		
Las conexiones entre las ideas son claras y adecuadas.		
Capacidad de argumentación		
Los argumentos que se presentan sustentan la tesis del ensayo.		
Los argumentos que se presentan son propios y sustentados teóricamente en fuentes de información confiables.		
Habilidades de comunicación escrita		
Las ideas se expresan de manera coherente y lógica.		
Las ideas están escritas correctamente y de manera clara.		

Las transiciones entre los párrafos son adecuadas.		
El escrito presenta una ortografía correcta.		
El vocabulario empleado en el desarrollo del ensayo es adecuado para la audiencia a la que está dirigido.		
Conclusión		
La conclusión presentada es original y sintetiza los argumentos presentados.		
Fuentes de información		
Las fuentes de información consultadas son confiables.		
Las fuentes de información son relevantes y variadas.		
Las fuentes de información son actuales y se encuentran relacionadas con el tema.		
Las fuentes de información están citadas correctamente de acuerdo con el formato especificado (APA, Chicago, entre otras).		

Fuente: Elaboración propia

RECURSOS EN LÍNEA

(Herramientas de evaluación a distancia que puedan apoyar al docente para generar y aplicar el instrumento).

Actualmente los alumnos y los profesores pueden apoyarse en aplicaciones tecnológicas a la hora de elaborar y revisar un ensayo. A continuación, se presenta una tabla con algunas aplicaciones que pueden ser de utilidad:

Tabla 3. Recursos y herramientas en línea para el ensayo

Recurso	Descripción
Writer plus https://play.google.com/store/apps/	Aplicación de escritura de notas, novelas, poemas y ensayos que permite al usuario crear y mantener sus archivos de manera sencilla y simple.
Grammarly https://www.grammarly.com	Aplicación que revisa la redacción de un texto de manera gratuita y sencilla.
Evernote https://evernote.com	Aplicación cuyo objetivo es la organización de información personal mediante archivos de notas.
iA Writer https://ia.net/es/writer	Aplicación que ofrece una experiencia de escritura única que permite la concentración en lo que se quiere decir y la creación de un texto estructurado y pulido.
Duplichecker https://www.duplichecker.com/es	Página electrónica para detectar plagio con un límite de 1,000 palabras, fácil de usar y con una versión gratuita.

Plagiarism Detector https://plagiarismdetector.net/es	Software gratuito de verificación de ensayos, con un límite de 1,000 palabras y que cuenta también con versión gratuita
Simil www.corpus.unam.mx/simil/default/corpus?id=707 https://www.dgcs.unam.mx/boletin/bdboletin/2017_306.html	Prototipo de la UNAM que detecta la paráfrasis o similitud total de oraciones en diferentes textos, y los ordena y categoriza de mayor a menor, pero no dictamina si se viola el derecho de autor, y tampoco señala si se citó o no la fuente del texto original.
Project Essay Grade https://www.measurementinc.com/products-services/automated-essay-scoring	Sistema de inteligencia artificial que utiliza técnicas estadísticas avanzadas para analizar la prosa escrita e identificar características intrínsecas de la escritura como fluidez, dicción, gramática, entre otras.

Existe una gran variedad de recursos tecnológicos que apoyan la elaboración y revisión de los ensayos; sin embargo, es importante que el profesor y los alumnos los utilicen como medios auxiliares, que tengan presente en todo momento cuál es el aprendizaje esperado por el cual se solicitó la elaboración del escrito y que, además, los recursos tecnológicos utilizados se complementen con los criterios descritos en los instrumentos de evaluación (rúbrica, lista de cotejo, entre otros).

CONCLUSIONES Y REFLEXIONES FINALES

- El ensayo es una herramienta que permite la evaluación de habilidades de pensamiento complejo y el desarrollo de la competencia comunicativa de escritura.
- Permite evaluar las capacidades de organización y síntesis de información, así como la argumentación por parte de los alumnos.
- El profesor puede emplear el ensayo como un recurso de evaluación formativa o sumativa.
- El ensayo se puede solicitar en las diferentes modalidades educativas: presencial, a distancia o híbrida.
- El profesor debe definir los criterios de evaluación previamente, con el objetivo de guiar la elaboración del ensayo y que el alumno conozca lo que se espera que domine.
- Los criterios de evaluación del ensayo se deben alinear al aprendizaje esperado que se requiere por parte del alumno.
- Se recomienda emplear en la evaluación del y para el aprendizaje de grupos pequeños, debido a que requiere una inversión de tiempo considerable para su calificación.
- Se sugiere el uso de otros instrumentos de evaluación como la rúbrica o lista de cotejo para que la evaluación del ensayo se lleve a cabo de manera sistemática.

A manera de síntesis revise el video "Capítulo Ensayo" Youtube. https://youtu.be/q_pbDWLvSkE

REFERENCIAS

Anguiano-López, M., Huerta-Rodríguez, J., Ibarra-Sepúlveda, J. y Almazán-Olachea, K. (2014). *Manual básico para la escritura de ensayos. Serie: Educación e Innovación (LEI)*. http://ideamex.com.mx/sites/default/files/pdf/manual_basico_para_escritura.pdf

Barrera, J. A., Castillo, M. (2019). *La competencia escrita, oral y gráfica en el currículo educativo*. https://idus.us.es/bitstream/handle/11441/99994/ase31_art05_barrera-vera_castillo-garc%C3%ADa%20%281%29.pdf?sequence=1&isAllowed=y

Díaz, D. (2004). *Cómo se elabora un ensayo*. *Acción Pedagógica, 13*(1), 108-113. http://www.saber.ula.ve/bitstream/123456789/17148/2/documento2.pdf

Errázuriz, María Constanza (2019). *Desempeño escrito de estudiantes de programas de formación inicial docente: ¿Cómo es la calidad del proceso de escritura de sus ensayos?* Lengua y Habla, (23),224-242. https://www.redalyc.org/articulo.oa?id=511966657012

Mazer, J. (2012). *Application essays as an effective tool for assessing instruction in the basic Communication course: A follow-up study. Journal of the Scholarship of Teaching and Learning, 12*(4). https://bit.ly/2XfOy6B

Muñoz, Carla, & Valenzuela, Jorge. (2015). *Psychometric Features of a Rubric to Assess Academic Writing at University Level*. *Formación universitaria, 8*(6), 75-84. https://dx.doi.org/10.4067/S0718-50062015000600010

Montoya, A. y Pérez, A. [María Azucena Montoya Magno] (29 de octubre de 2021). *Capítulo Ensayo* [Video]. Youtube. https://youtu.be/q_pbDWLvSkE

Montoya, A. y Pérez, A. [María Azucena Montoya Magno] (5 de mayo de 2022). Ensayo [Video]. Youtube. https://youtu.be/CMNtfZ9ub3Y

Ortega, E. (2021). Evaluación para el aprendizaje: enfoque cualitativo. Ensayo y tipos de ensayo [MOOC] https://www.coursera.org/lecture/evaluacion-cualitativa/ensayo-y-tipos-de-ensayo-CbMSI

Osorio, Elisa Bigi y Romero, Marisol García y Guerrero, Edixon Chacón (2019) ¿Qué textos académicos escriben los estudiantes universitarios de Educación? Zona Próxima, (31),25-55. https://www.redalyc.org/articulo.oa?id=85362906003

Ruiz-Terroba, R., Vázquez, E. y Sevillano-García, M. (2017). *La rúbrica de evaluación de la competencia en expresión escrita. Percepción del alumnado sobre su funcionalidad*. https://www.redalyc.org/pdf/2591/259153707010.pdf

Capítulo 27

ENSAYO RESTRINGIDO

Ana Laura Pérez Díaz, María Azucena Montoya Magno

"El arte de escribir consiste en decir mucho con pocas palabras."
Antón Chejóv

INTRODUCCIÓN

En el ámbito universitario es frecuente la solicitud de textos académicos escritos por los propios estudiantes, la mayoría de ellos con fines evaluativos para demostrar los conocimientos adquiridos, el dominio de algún tema en específico, el desarrollo de habilidades expresivas y comunicativas, entre otros.

Bajo esta premisa, las prácticas de lectura y escritura son indispensables para que los estudiantes se adentren en el conocimiento de las características de los textos solicitados en el contexto universitario (Osorio et al. 2019).

Durante el proceso enseñanza-aprendizaje, los profesores buscan que sus alumnos sean capaces de comprender lo que se les enseña en el curso y que no solo memoricen la información. El término comprender debe abarcar aspectos tales como la capacidad de análisis, síntesis y argumentación. En este sentido, una de las herramientas que permitirá al profesor evaluar con precisión lo que quiere que los alumnos hayan comprendido en un determinado momento del curso es el ensayo restringido. Si el profesor desea evaluar la capacidad del alumno para expresar, integrar y organizar ideas puede hacer uso del ensayo restringido, que consiste en una pregunta o un conjunto de preguntas, en las que el alumno elabora su respuesta.

En el ensayo restringido es el profesor quien determina qué tipo de respuesta espera por parte del alumno, y dependerá del aprendizaje que se quiera evaluar. El propósito de este capítulo es poner al alcance de los profesores una herramienta que los ayude en el diseño y uso de ensayos restringidos en la evaluación del y para el aprendizaje.

Este capítulo está organizado en seis secciones: en la primera se explica qué es, en la segunda se detallan los pasos para diseñarlo, en la tercera se recomienda cómo aplicarlo, además de que en esta sección se presenta información del desarrollo del ensayo restringido en un marco de la evaluación a distancia, para ello se especifica cómo diseñarlo y aplicarlo en

dicha modalidad. En la cuarta sección se puntualiza cómo analizar sus resultados (ya sea en modalidad presencial, a distancia o híbrida). La quinta sección presenta ejemplos y recursos en línea y, por último, en la sexta se presentan conclusiones y reflexiones finales.

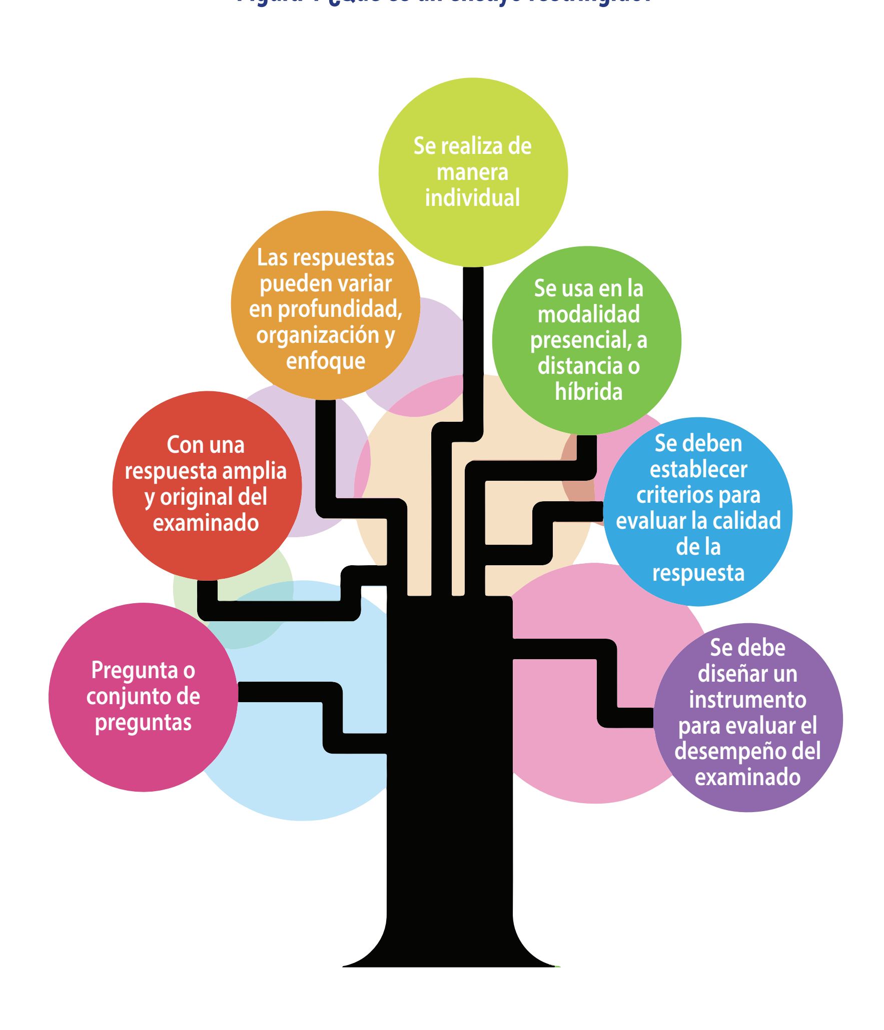
Este capítulo es un recurso de autoestudio, reflexión y consulta que puede leerse o no en orden. Pretende ser un punto de partida para que el profesor explore las posibilidades que ofrecen los ensayos restringidos, reflexione sobre sus prácticas de evaluación e intercambie ideas con otros colegas.

¿QUÉ ES?

- El ensayo restringido consiste en una pregunta o conjunto de preguntas respecto de un tema, que requieren una respuesta amplia y original por parte del alumno.
- También se le conoce como prueba de ensayo o prueba de composición.
- Se caracteriza porque las respuestas a la pregunta o conjunto de preguntas son construidas por los alumnos de acuerdo con el contenido que se indica se espera de ellos, por lo que no hay respuestas correctas únicas. Las respuestas pueden variar en cuanto a extensión, profundidad, organización de la información, enfoque, etcétera, y ser igualmente correctas. (Basabe et al., 2020)
- Generalmente este tipo de prueba es realizada de manera individual y puede aplicarse en las distintas modalidades educativas: presencial, a distancia o híbrida.
- El profesor deberá indicar a los alumnos cuáles son los criterios para evaluar la calidad de la respuesta y el instrumento que utilizará para valorar las respuestas (INEE, 2020).
- El ensayo restringido puede utilizarse en diferentes niveles educativos, aunque generalmente su uso se recomienda en los niveles medio superior y superior, en los cuales se requiere evaluar niveles complejos de razonamiento.
- Es útil para conocer el grado en que se han logrado ciertos objetivos de aprendizaje que involucran la habilidad para organizar, sintetizar, estructurar e integrar ideas y comunicar información, el razonamiento lógico y la resolución de problemas. (Ministerio de Educación y formación profesional et al., 2020)
- Una de las ventajas del uso del ensayo restringido, es que al construir el alumno su respuesta, evita la adivinación o socialización de la respuesta correcta. Además, permite a los profesores detectar problemas de razonamiento en los alumnos a partir de las respuestas dadas por estos. Otra ventaja es que permite obtener información sobre el desarrollo de procesos mentales complejos que no puede obtenerse mediante otros instrumentos de evaluación, además de que el profesor cuenta con mayores elementos para dar *feedback* o retroalimentación y abrir la posibilidad de diálogo con los estudiantes (González, 2002). Una de las desventajas al utilizar el ensayo restringido como un recurso de evaluación, es que solo se puede evaluar una muestra limitada de contenido debido a la cantidad de tiempo que emplean los alumnos para responder las preguntas; y la calificación de las respuestas requiere una inversión de tiempo considerable por parte del profesor.

- Otra de las desventajas en el empleo del ensayo restringido, es que puede presentar problemas de validez, debido a que se involucra la subjetividad en el procedimiento de calificación. Pues, aunque se cuente con lineamientos para realizar este procedimiento, también influyen los aspectos emocionales y de tiempo del calificador al momento de llevarla a cabo, la relación personal o impresión que se tenga del alumno, así como la presentación del material escrito.

Figura 1 ¿Qué es un ensayo restringido?



¿CÓMO LO DISEÑO?

Para el diseño del ensayo restringido es necesario llevar a cabo una serie de pasos sistemáticos enfocados al cumplimiento del objetivo de la evaluación. Para ello, es importante que exista una alineación entre la enseñanza, el aprendizaje y la evaluación, por lo que se recomienda se analice el programa de la asignatura o curso, identificando los temas y aprendizajes esperados que se evaluarán con este instrumento.

Planeación

1) **Seleccione los aprendizajes esperados que se evaluarán por medio del ensayo restringido.**
   ¿El ensayo restringido es el mejor instrumento para valorar ese aprendizaje esperado o conjunto de aprendizajes que debe dominar el alumno?
2) **Determine el tipo de evaluación.**
   El ensayo restringido suele utilizarse en la evaluación sumativa, ya que permite identificar el logro de los aprendizajes esperados al término de un tema, una unidad o un curso. Aunque también es muy útil en la evaluación formativa, puesto que permitirá al profesor conocer el grado de comprensión que poseen los alumnos respecto a un tema y orientarlos respecto a los conocimientos o habilidades que pueden mejorar en aras del logro de los aprendizajes esperados.
3) **Especifique quiénes participarán en la evaluación.**
   El profesor es el principal actor en la evaluación del ensayo restringido (heteroevaluación), pues es quien establece los criterios de valoración para cada una de las respuestas y el responsable de proporcionar la retroalimentación a los alumnos.
4) **Redacte la pregunta o conjunto de preguntas de manera clara y precisa, para que la tarea quede claramente definida y sin ambigüedades.**
   Utilice verbos directivos como: explique, justifique, proponga, defienda o evalúe.
   Las preguntas pueden aplicarse de manera presencial, a distancia o híbrida.
5) **Establezca los elementos o criterios que debe cubrir la respuesta o respuestas esperadas.**
   Formule la guía de la respuesta modelo, así como los elementos que deberá cubrir para considerarse correcta, y comunique dicha información a los alumnos con el objetivo de que conozcan lo que se espera de ellos.
6) **Solicite apoyo a un colega para que revise la o las preguntas del ensayo restringido y la o las respuestas esperadas.**
   ¿Están alineados el aprendizaje esperado, la pregunta de ensayo restringido y la respuesta esperada? Defina las características del ensayo restringido.
   Establezca el número de preguntas y extensión de las respuestas esperadas.
7) **Determine los instrumentos que acompañarán la evaluación del ensayo restringido**
   Para que la evaluación del ensayo restringido se lleve a cabo de una manera sistemática, se recomienda el uso de instrumentos tales como la rúbrica, la lista de cotejo y la escala de calificación. Se recomienda que para ello se dirija a los capítulos 14 y 15 de este libro.

Figura 2. Planeación del ensayo restringido

Planeación del Ensayo restringido

1 Seleccione los aprendizajes esperados que se evaluarán por medio del ensayo restringido

Determine el tipo de evaluación: formativa o sumativa 2

3 Especifique quiénes participarán en la evaluación

Redacte la pregunta o conjunto de preguntas de manera clara y precisa 4

5 Establezca los elementos o criterios que debe cubrir la respuesta o respuestas esperadas

Solicite apoyo a un colega para que revise la o las preguntas del ensayo restringido y la o las respuestas esperadas 6

7 Defina las características del ensayo restringido

Determine los instrumentos que acompañarán la evaluación del ensayo restringido 8

DISEÑO

1) **Se sugiere iniciar la redacción de las preguntas de ensayo restringido con verbos directivos, por ejemplo: compare, ejemplifique, distinga, entre otros.**
   El uso de estos verbos guía a los estudiantes respecto a lo que se espera que respondan.
2) **Se recomienda redactar las preguntas del ensayo restringido de manera clara y precisa.**
3) **Recuerde que el objetivo del ensayo restringido es evaluar qué tan bien responde el alumno a una instrucción específica, si en las preguntas existe la posibilidad de una mala interpretación no se cumpliría el objetivo de la misma. Incluya una sección de datos personales, las instrucciones, las preguntas y los criterios de evaluación en el ensayo restringido.**
4) **Edite la versión final del documento.**

El documento que genere puede ser entregado a los alumnos de manera impresa o digital, dependiendo de la modalidad en la que sea implementada y puede ser presencial, a distancia o híbrida. Si existiera alguna duda sobre los criterios de evaluación se recomienda que se aclare con los alumnos.

Revisión

- Se recomienda que un colega revise la o las preguntas de ensayo restringido antes de utilizarlas con los alumnos, a fin de determinar si están alineadas con el aprendizaje esperado y con la respuesta esperada.
- Se sugiere contar con suficientes preguntas de ensayo restringido que permitan modificar periódicamente el instrumento de evaluación.
- Se deben anotar las incidencias que puedan presentarse cuando se utilice este instrumento de evaluación.

¿CÓMO LO APLICO?

- El ensayo restringido, como recurso de evaluación, puede utilizarse cuando sea necesario que los alumnos demuestren el conocimiento, comprensión o aplicación de determinada información o contenido.
- Se puede usar cuando se requiera evaluar la habilidad de los alumnos para sintetizar información y comunicar ideas o contenidos de manera clara y precisa.
- Se puede aplicar cuando se necesite evaluar la capacidad de argumentación en los alumnos.
- Se recomienda que sea realizado de manera individual, dentro o fuera del espacio educativo, pues ello permitirá al profesor conocer el grado en que cada alumno ha desarrollado distintas habilidades de pensamiento complejo, como las mencionadas en los puntos anteriores.

- Puede aplicarse al finalizar una unidad o un curso, apoyando a la evaluación formativa o sumativa para identificar las fortalezas y áreas de oportunidad de los alumnos respecto a los aprendizajes esperados con el propósito de llevar a cabo acciones educativas que permitan el logro de estos. Es necesario que, en todo momento, se tengan presentes los criterios que se considerarán en la evaluación del ensayo, con la finalidad de evitar, en la medida de lo posible, juicios subjetivos durante el proceso de evaluación.

El ensayo restringido en evaluación a distancia

El ensayo restringido puede implementarse en la modalidad presencial, a distancia o híbrida de manera síncrona y asíncrona. Basabe et al. (2020) mencionan que, de manera síncrona, los alumnos pueden responder un ensayo restringido durante un determinado lapso de tiempo, apoyados de una plataforma virtual como Zoom, Cisco, Webex Meetings, Skype, Google Hangouts, Google Meet, entre otros; y de forma asíncrona, los estudiantes pueden realizar esta prueba de manera diferida en el tiempo, pero con una fecha de entrega establecida. Tanto en la modalidad a distancia como presencial, el docente debe considerar la retroalimentación individual de las preguntas de acuerdo con la rúbrica de evaluación y puede realizarse directamente en la misma herramienta virtual donde se realizó la prueba de ensayo (Pontificia Universidad Católica de Valparaíso, 2020.). Lo más importante de proporcionar retroalimentación a los alumnos es identificar las fortalezas y las áreas de oportunidad en las cuales deberá trabajar para lograr los aprendizajes esperados.

Diseño de la rúbrica de calificación para el ensayo restringido en educación a distancia

Si el ensayo restringido es considerado como un recurso de evaluación individual, es necesario realizar la elaboración de una rúbrica para evaluarlo. Para ello, es fundamental que se diseñe especificando a los alumnos los niveles de logro que se desean alcanzar, definiendo de manera clara y precisa los criterios que serán evaluados junto con sus descripciones. Es muy importante que durante el diseño de la rúbrica especifique los atributos que se desean evaluar y que estos estén alineados al aprendizaje o aprendizajes a evaluar a través del ensayo restringido.

¿Cómo lo diseño a distancia?

Independientemente que las pruebas de ensayo o ensayo restringido sean aplicadas por el profesor en modalidad presencial, a distancia o híbrida, es necesario considerar una serie de pasos sistemáticos para su planeación y desarrollo, de manera que exista una alineación entre aquello que aprenden los alumnos y los objetivos de aprendizaje que se desean evaluar.

En entornos virtuales existen múltiples herramientas que permiten diseñar de manera gratuita pruebas de ensayo que sean parte de la evaluación sumativa o formativa de los alumnos. Se sugiere que indague y experimente para saber cuál de estas herramientas es la que más se ajusta a sus necesidades de evaluación.

Según Basabe et al. (2020), para elegir la herramienta o plataforma más adecuada es conveniente verificar que las herramientas en las que se diseñe el ensayo restringido faciliten el envío de las respuestas de los estudiantes, pueda configurar la fecha de entrega y cuente con

un apartado que le permita al profesor proporcionar la realimentación al alumno, esta última es una actividad fundamental de cualquier proceso de evaluación.

Es importante resaltar que el hecho de contar con una plataforma virtual para el envío y entrega de los productos de los alumnos permite que el profesor tenga un registro sistemático de los trabajos entregados por cada uno de ellos.

Con el objetivo de evitar problemas técnicos que pudieran presentarse durante el envío de los productos realizados por los alumnos, se recomienda preparar algún breve tutorial sobre la plataforma virtual en la que los alumnos resolverán la prueba de ensayo para facilitar su uso a quienes tengan poca experiencia en el manejo de estas herramientas digitales (Basabe et al., 2020).

¿Cómo lo instrumento?

- El ensayo restringido es una técnica de evaluación que puede llevarse a cabo en las modalidades presencial, a distancia o híbrida.
- Para la elaboración de la prueba de ensayo es necesario que el profesor determine los resultados de aprendizaje que desea evaluar, con el objetivo de guiar el diseño de la prueba.
- Como recurso de evaluación, puede utilizarse cuando sea necesario que los alumnos demuestren el conocimiento, comprensión o aplicación de determinada información o contenido.
- Se puede usar cuando se requiera evaluar la habilidad de los alumnos para sintetizar información y comunicar ideas o contenidos de manera clara y precisa.
- Se puede aplicar cuando se necesite evaluar la capacidad de argumentación en los alumnos.
- Se recomienda que sea realizado de manera individual, dentro o fuera del espacio educativo, pues ello permitirá al profesor conocer el grado en que cada alumno ha desarrollado distintas habilidades de pensamiento complejo, como las mencionadas en los puntos anteriores.
- En caso de que el ensayo restringido se lleve a cabo en la modalidad en línea, es necesario que el profesor comunique a los alumnos en la plataforma virtual la rúbrica en la cual se especifiquen los criterios a evaluar, los niveles de desempeño en cada caso y la fecha límite de entrega.
- Puede aplicarse al finalizar una unidad o un curso, apoyando a la evaluación formativa o sumativa para identificar las fortalezas y áreas de oportunidad de los alumnos respecto a los aprendizajes esperados con el propósito de llevar a cabo acciones educativas que permitan el logro de estos. Es necesario que, en todo momento, se tengan presentes los criterios que se considerarán en la evaluación del ensayo, con la finalidad de evitar, en la medida de lo posible, juicios subjetivos durante el proceso de evaluación.

Consulte el video “Pasos para diseñar un ensayo restringido”
(Youtube. https://youtu.be/qyMW882ZtY8)

¿CÓMO ANALIZO SUS RESULTADOS?

- La pregunta o preguntas que integran el ensayo restringido se diseñan a partir de un marco de referencia que responde a un aprendizaje esperado dentro de una asignatura en particular. Es recomendable que cuando se analicen sus resultados no se pierdan de vista estos aspectos y se valore qué tanto se consolidó el aprendizaje de los alumnos, qué aspectos están poco consolidados y cuáles se han de trabajar en el espacio educativo.
- En caso de que el docente haya empleado una rúbrica analítica para evaluar las respuestas de los alumnos a las cuales se les haya asignado un puntaje, es necesario que realice la suma de cada uno los puntajes que haya asignado a cada criterio o elemento de evaluación, con el objetivo de obtener un puntaje total y de esta manera el nivel de desempeño del alumno.
- Se recomienda no limitar a una escala numérica los resultados obtenidos con el uso de un ensayo restringido, ya que el valor del uso de este tipo de instrumentos radica en la riqueza de información que puede proporcionar al profesor. Es indispensable que el profesor comunique a los alumnos todos los criterios a considerar en la evaluación de un ensayo restringido, independientemente de la modalidad (presencial, en línea o híbrida) y de la forma síncrona o asíncrona, para que al momento de analizar los resultados se tenga presente en qué medida se cumplieron cada una de las condiciones establecidas previamente por el profesor. La utilidad de este tipo de pruebas se incrementa cuando el alumno recibe retroalimentación oportuna y fundamentada sobre sus respuestas.

EJEMPLOS Y RECURSOS EN LÍNEA

Ejemplos

Licenciatura en Psicología
Asignatura: Aprendizaje y conducta adaptativa III

Objetivo general:
El alumno comprenderá la conducta de los organismos como resultado de mecanismos que favorezcan la adaptación.
El alumno conocerá los principios básicos de acción y el papel que juegan en la explicación y modificación del comportamiento humano.

Objetivo específico:
Proporcionar al alumno, por medio de prácticas, condiciones para la consolidación de conceptos y su relación con el ambiente experimental y natural.

Pregunta de un ensayo restringido
Pavlov observó que los perros condicionados por él salivaban ante el sonido de una

campana. ¿Cómo explicaría esto? Su respuesta deberá ser de aproximadamente de media cuartilla de extensión.

Respuesta esperada
Pavlov se propuso enseñarles a los perros a salivar cuando no estuviera presente la comida. Para ello, diseñó un experimento en el cual hacía sonar una campana poco antes de traer la comida al cuarto donde se encontraban los perros. El sonido de una campana, normalmente, no hace salivar a los perros, pero después de oírla varias veces poco antes de ser alimentados, los animales, comenzaban a salivar en cuanto sonaba la campana. Era como si hubieran aprendido que dicha campana indicaba la presencia de comida, entonces su boca empezaba a salivar, aunque no les presentarán comida. Se les había condicionado a salivar ante un nuevo estimulo: la campana que normalmente no provocaba esta respuesta.

Tabla 1. Rúbrica para evaluar el ensayo restringido.

Instrucciones: A continuación, asigne un puntaje a la respuesta presentada por el alumno considerando su desempeño en el aspecto que se requiere evaluar.

Aspectos a evaluar en la prueba de ensayo restringido	Excelente (3 puntos)	Satisfactorio (2 puntos)	Suficiente (1 punto)	Deficiente (0 puntos)	Puntaje asignado
TEMA A DESARROLLAR					
La respuesta hace referencia a los elementos del condicionamiento clásico de Pavlov.	Hace referencia a los cuatro conceptos básicos del condicionamiento clásico: Estímulo incondicionado (EI-comida), Respuesta incondicionada (RI-salivar), Estímulo condicionado (EC-campana), Respuesta condicionada (RC-salivar).	Hace referencia a tres de los siguientes conceptos básicos del condicionamiento clásico: Estímulo incondicionado (EI-comida), Respuesta incondicionada (RI-salivar), Estímulo condicionado (EC-campana), Respuesta condicionada (RC-salivar).	Hace referencia a dos de los siguientes conceptos básicos del condicionamiento clásico: Estímulo incondicionado (EI-comida), Respuesta incondicionada (RI-salivar), Estímulo condicionado (EC-campana), Respuesta condicionada (RC-salivar).	Hace referencia a uno o ninguno de los siguientes conceptos básicos del condicionamiento clásico: Estímulo incondicionado (EI-comida), Respuesta incondicionada (RI-salivar), Estímulo condicionado (EC-campana), Respuesta condicionada (RC-salivar).	

HABILIDADES					
Capacidad de análisis y síntesis de la información	Presenta su respuesta de manera ordenada, lógica y secuencial, construye un punto de vista a partir de sus conocimientos. En su respuesta la información se presenta de manera sintética, precisa y clara.	Presenta su respuesta de manera ordenada, lógica y secuencial. En su respuesta la información se presenta de manera sintética, precisa y clara.	Presenta su respuesta de manera ordenada. En su respuesta la información se presenta de manera clara.	Presenta su respuesta de manera ordenada. Sin embargo, en su respuesta la información se presenta de manera ambigua.	

Fuente: Elaboración propia

Tabla 2. Lista de cotejo para la evaluación de un ensayo restringido

Esta lista tiene como propósito realizar una valoración sistemática de un ensayo restringido.

Datos de identificación

Nombre del alumno:	
Tema:	
Objetivo de aprendizaje:	

Instrucciones: A continuación, se presentan una serie de elementos que deberá contener el ensayo restringido para su evaluación. Por favor marque la casilla que indique si el ensayo restringido cumple o no con el elemento mencionado.

Elemento a valorar	**Cumple**	
	Sí	No
El tema a desarrollar está formulado de manera clara y precisa.		
Se identifica de manera clara la tesis sobre la que se sustentará el tema a desarrollar.		
Existe una articulación adecuada entre los párrafos de la respuesta.		
Existe dominio del contenido por parte del alumno.		
Las ideas presentan una estructura clara, coherente, fluida y lógica.		
Las ideas presentadas son relevantes.		

Las ideas que se presentan son propias y sustentadas teóricamente en fuentes de información confiables.		
Las conexiones entre las partes que integran el ensayo restringido son precisas, adecuadas y claras.		
Las ideas están escritas correctamente siguiendo las reglas de gramática y sintaxis.		
El texto presenta una ortografía correcta.		
El lenguaje empleado es amplio y adecuado.		
La conclusión es original y presenta inferencias sobre el tema desarrollado.		

Fuente: Elaboración propia

RECURSOS EN LÍNEA

(Herramientas de evaluación a distancia que puedan apoyar al docente para generar y aplicar el instrumento).

En el contexto de la pandemia que se suscita a nivel mundial, los profesores se han visto en la necesidad de adaptar sus estrategias de enseñanza, aprendizaje y de evaluación de modelos presenciales a modelos a distancia o híbridos. Por esta razón, se ha favorecido el uso de los recursos tecnológicos y de las plataformas digitales que coadyuven en los procesos de enseñanza, aprendizaje y evaluación, siempre buscando que dichos procesos se encuentren alineados con los objetivos de aprendizaje esperados en los alumnos.

En el caso del ensayo restringido o pruebas de ensayo existen múltiples recursos los cuales puede emplear el profesor para diseñar este instrumento que coadyuve en sus procesos de evaluación, ya sea de tipo formativa o sumativa, los cuales se enlistan a continuación:

Tabla 3. Recursos y herramientas en línea para el ensayo restringido

Recurso	Descripción
Moodle https://moodle.org/?lang=es	Es una plataforma en el cual el docente puede incluir y almacenar las preguntas tipo de ensayo de acuerdo con los aprendizajes que desea evaluar en el alumno. Adicionalmente, aparece un campo donde se puede indicar el puntaje asignado a cada pregunta y un espacio para proporcionar la retroalimentación al alumno. Las respuestas se califican de manera manual, es decir, no es automatizada. Es un recurso de uso gratuito.
iGiveTest https://igivetest.com/es/	Es herramienta que permite crear pruebas en línea utilizando preguntas de opción múltiple, verdadero o falso y de respuesta abierta. El docente puede asignar la puntuación a cada una de ellas y enviar sus comentarios de retroalimentación a las respuestas. Es un recurso que cuenta con un periodo de prueba gratuito y en caso de que sea funcional para las necesidades del docente puede continuar usándolo, seleccionando uno de los planes de pago.

Recurso	Descripción
Online Exam Builder https://www.onlineexambuilder.com/es/	Es una plataforma permite generar una prueba con diversos tipos de preguntas como preguntas de texto libre, de opción múltiple, de audio y video, así como establecer un límite de tiempo para responder cada pregunta o el examen en su totalidad. Puede determinarse la fecha de inicio y finalización de la prueba y proporcionar retroalimentación a los alumnos.
QuestBase https://questbase.com/	Es otra herramienta que permite diseñar diversas pruebas en línea y realizar correcciones de manera automática. Dentro de sus características fundamentales es que permite incluir diversos tipos de preguntas, texto libre, falso-verdadero, opción múltiple y el docente puede especificar el tiempo límite para la realización de la prueba. Es un recurso que puede utilizarse de manera gratuita con varias restricciones respecto al número de pruebas a diseñar y los datos obtenidos de la aplicación de la misma. En caso de que el recurso sea funcional para las necesidades del docente puede continuar usándolo, seleccionando uno de los planes de pago.

Es importante enfatizar que, si bien es cierto que el uso de herramientas digitales es diverso y novedoso, y además facilita los procesos de evaluación de los alumnos, es fundamental no perder de vista que lo más importante es un buen diseño de un instrumento de evaluación a través del cual se obtenga información válida y confiable del desempeño del alumno.

CONCLUSIONES Y REFLEXIONES FINALES

- El ensayo restringido es una herramienta que permite evaluar la construcción de respuestas originales por parte de los alumnos a partir de preguntas y criterios de respuesta previamente establecidos por el profesor.
- La evaluación de las respuestas de los alumnos puede realizarse a partir de respuestas modelo construidas por el profesor. No hay respuestas únicas, pero si se apegan a los criterios establecidos por el profesor pueden ser correctas.
- Como el alumno es quien construye su respuesta, se evita la adivinación o socialización de la respuesta correcta.
- Es útil para evaluar habilidades de nivel superior como capacidad de organización, de síntesis, de integración de ideas, el razonamiento crítico y lógico.
- Esta herramienta permite evaluar contenidos temáticos limitados debido a que el tiempo de construcción y evaluación de las preguntas requiere una inversión de tiempo considerable por parte del profesor.
- Se recomienda que las preguntas y las respuestas modelo sean revisadas por otros expertos en el tema, con la finalidad de que verifiquen su pertinencia y en caso de ser necesario, el profesor realice los ajustes necesarios con el objetivo de contar con un buen instrumento de evaluación.

- Se sugiere el uso de otros instrumentos de evaluación como la rúbrica para que la evaluación del ensayo restringido se lleve a cabo de manera sistemática.
- Se sugiere el uso de la rúbrica analítica para la valoración de las respuestas de los alumnos, en la cual se especifiquen los criterios o elementos a evaluar y el profesor puede asignar un puntaje en función de los criterios que debe cumplir el alumno al momento de dar su respuesta.
- Además, este tipo de rúbrica es una guía que puede orientar al alumno al momento de elaborar su respuesta, ya que se especifica de manera clara que es lo que se desea evaluar o lo que se espera que él responda. El profesor puede hacer uso de múltiples recursos tecnológicos o plataformas digitales diversas y novedosas que le permiten diseñar las pruebas de ensayo y facilitar sus procesos de evaluación, sin embargo, siempre debe tener presente que, si bien son importantes los recursos empleados, la prioridad es tener un buen diseño del instrumento que le permita obtener información válida y confiable del desempeño de los alumnos.
- El ensayo restringido considerado como un instrumento de evaluación formativa es muy útil para identificar las fortalezas y las áreas de oportunidad en los contenidos y habilidades que los alumnos deben mejorar.

A manera de resumen se sugiere revisar el video "Capítulo Ensayo restringido"(Youtube. https://youtu.be/d5jbF0R-CG4).

REFERENCIAS

Basabe, L., Leal Falduti, B. y Tornese, D. (2020). *Diseño de exámenes con ítems de respuesta abierta*. Citep. Centro de Innovación en Tecnología y Pedagogía. http://citep.rec.uba.ar/wp-content/uploads/2020/05/AcaDocs_D09_Dise%C3%B1o-de-ex%C3%A1menes-escritos-con-%C3%ADtems-de-respuesta-abierta-1.pdf

Educational Evaluation. (2011). *Essay test*. http://educationalevaluationhema.blogspot.com/2011/03/essay-test.html

González, H. (2002). *Tipos de preguntas: ¿Cerradas o Abiertas?* http://eduteka.icesi.edu.co/articulos/PruebasAbiertasCerradas

Innovación y desarrollo docente, (2020). *10 herramientas online para crear y corregir exámenes.* https://iddocente.com/herramientas-online-crear-corregir-examenes/

Instituto Nacional para la Evaluación de la Educación. (2020). *Cuadernillo técnico de evaluación educativa. Desarrollo de instrumentos de evaluación: pruebas.* https://www.inee.edu.mx/wp-content/uploads/2019/08/P2A354.pdf

Lara-Inostroza, F. y Cabrera-Pommiez, M. (2015). *Guía de evaluación educativa UDLA*. Santiago de Chile, Chile: UDLA. https://www.academia.edu/27251762/Gu%C3%ADa_de_evaluaci%C3%B3n_educativa

Ministerio De Educación y Formación Profesional, Secretaría de Estado de Educación, Dirección General de Evaluación y Cooperación Territorial e Instituto Nacional de Evaluación Educativa.

(2020). *Guía para la elaboración de pruebas de evaluación educativa.* https://redined.mecd.gob.es/xmlui/bitstream/handle/11162/202348/GUIA.pdf?sequence=1&isAllowed=y

Ortega, E. (2021). *Evaluación para el aprendizaje: enfoque cualitativo. Ensayo y tipos de ensayo* [MOOC] https://www.coursera.org/lecture/evaluacion-cualitativa/ensayo-y-tipos-de-ensayo-CbMSI

Pontificia Universidad Católica de Valparaíso. (2020). *La Prueba de Respuesta Abierta Extensa como instrumento de evaluación.* http://innovaciondocente.ucv.cl/wp-content/uploads/2020/11/Fichas_evaluacion_virtual_Prueba_Extensa.pdf

Pérez, A. y Montoya, A. [Ana Laura Pérez Díaz] (09 de mayo de 2022). *Capítulo Ensayo restringido* [Video]. Youtube. https://youtu.be/d5jbF0R-CG4

Pérez, A. y Montoya, A. [Ana Laura Pérez Díaz] (09 de mayo de 2022). *Pasos para diseñar un ensayo restringido* [Video]. Youtube. https://youtu.be/qyMW882ZtY8

Capítulo 28

ESTUDIO DE CASO

Guadalupe Soto Estrada, Enrique Ricardo Buzo Casanova

"La evaluación deberá servir entonces, para reorientar y planificar la práctica educativa. Conocer lo que ocurre en el aula a partir de los procesos pedagógicos empleados y su incidencia en el aprendizaje del alumno, reorientando cuantas veces fuere necesario los procesos durante su desarrollo, es una de las funciones más importantes de la evaluación."

SHMIEDER, STOCKER Y TITONE

INTRODUCCIÓN

En la actualidad es fundamental conocer el proceso de aprendizaje de los alumnos, el cual puede valorarse mediante pruebas objetivas, así como con instrumentos, técnicas y estrategias cualitativas de evaluación, que permiten conocer los procesos cognitivos a lo largo del proceso enseñanza-aprendizaje en un curso y brindar retroalimentación a los alumnos. Dichos instrumentos de evaluación proporcionan al profesor la capacidad de extraer información valiosa que le permita la toma de decisiones respecto al proceso enseñanza-aprendizaje para mejorar los conocimientos y habilidades del alumnado.

El estudio de caso es una herramienta de evaluación cualitativa que le permite al alumno enfrentarse a situaciones que puede experimentar en la vida real y tomar una postura al respecto en un ambiente seguro. De ahí que una de las ventajas que ofrece esta herramienta es el entrenamiento de los alumnos en la elaboración de soluciones para los problemas o situaciones que se les plantean, con la ventaja de contar con la retroalimentación de su profesor, quien, al tener un bagaje teórico extenso y experiencia profesional, puede ampliar el panorama de los alumnos y exponer la forma en que él lo abordaría. Además, la discusión de grupo que genera el análisis de caso también brinda oportunidad para conocer la opinión, la postura, las estrategias de intervención y la actitud de los compañeros de clase, enriqueciendo la visión del alumno a partir de la interacción con sus pares y su profesor.

El objetivo de este texto es proporcionar a los docentes información útil sobre el estudio de caso como herramienta de evaluación cualitativa del aprendizaje, con la finalidad de utilizarla en el espacio educativo cuando se considere que es la mejor opción para evaluar el aprendizaje.

El capítulo está organizado en siete secciones. En la primera se explica qué es un estudio de caso, en la segunda se detallan los pasos para diseñar una evaluación cualitativa mediante

un estudio de caso, en la tercera se recomienda cómo aplicarlo de forma presencial y a distancia, en la cuarta se ve cómo analizarlo, en la quinta se muestran ejemplos, en la sexta se ofrecen recursos en línea para valorar estudios de caso y en la séptima se presentan algunas conclusiones y recomendaciones.

Este texto es un recurso de autoestudio, reflexión y consulta que puede leerse o no en orden. Pretende ser un punto de partida para que el profesor o facilitador explore las posibilidades que ofrece el estudio de caso, reflexione sobre sus prácticas de evaluación e intercambie ideas con otros colegas.

¿QUÉ ES?

- El estudio de caso es una herramienta para fines de evaluación del y para el aprendizaje que se utiliza para comprender en profundidad la realidad y tiene un amplio uso en diversas áreas de conocimiento.
- De acuerdo con Lawrence (1953, citado en Wasserman 1994): "Un buen caso es el vehículo por medio del cual se lleva al espacio educativo un trozo de realidad a fin de que los alumnos y el profesor lo examinen minuciosamente. Un buen estudio de caso mantiene centrada la discusión en alguno de los escenarios con los que uno debe enfrentarse en ciertas situaciones de la vida real".
- Mediante el uso del estudio de caso como herramienta para la evaluación para el aprendizaje se puede obtener información diversa; por ejemplo, sobre la manera en que los alumnos generan soluciones efectivas ante situaciones reales, estudiar problemas complejos, analizar la forma en que se deben interrelacionar distintos campos disciplinarios para comprender un caso, o implementar estrategias de intervención terapéutica, entre otras.
- Se recomienda que un estudio de caso sea:
  Particular: Que esté centrado en una situación, evento, programa o fenómeno específico, lo que lo hace ser una herramienta óptima en problemas prácticos de evaluación del y para el aprendizaje.
  Descriptivo: El producto del estudio de caso es una descripción rica del fenómeno bajo estudio.
  Heurístico o de descubrimiento: favorece la comprensión del fenómeno bajo estudio.
- Cuando el profesor o facilitador utilice esta herramienta para evaluar el resultado de aprendizaje de su interés, en función del caso que utilice, podrá obtener información sobre la manera en que el alumno usa su experiencia, analiza el caso y elabora interpretaciones para lograr su entendimiento y/o solución.
- Con la finalidad de aprovechar la riqueza de la información que el profesor puede obtener con el uso de esta herramienta, es preciso realizar una adecuada planeación y diseño, así como un análisis y empleo óptimo de los resultados.

¿CÓMO LO DISEÑO?

- Seleccionar un caso coherente con el tema y el objetivo de aprendizaje que se pretende valorar. El docente puede realizar algunas adecuaciones al contenido para que éste resulte comprensible para los alumnos y contenga información que favorezca el análisis del tema que se está estudiando. Éste es el caso de autores como Donoso-Vázquez y Sánchez-Martí (2014), Merriam (2009) y Stake (2007).
- Es indispensable cerciorarse de mantener el anonimato de las personas involucradas o bien, de solicitar su autorización para utilizarlo.

Planeación

Se sugiere que el profesor o facilitador tome como eje rector en la planeación de sus clases y del ciclo escolar el mapa curricular, el perfil de egreso y el programa de su asignatura, bloque o curso. En dichos documentos se establecen los conocimientos, habilidades, actitudes y valores que se pretende que el alumno adquiera y desarrolle. Con base en lo anterior, podrá identificar la herramienta, estrategia o instrumento más pertinente para evaluar el logro de los aprendizajes esperados.

1) **Identifique el o los objetivos de aprendizaje que quiere valorar.**
   ¿El estudio de caso es el mejor instrumento para valorar ese aprendizaje?
2) **Defina el tipo de evaluación para la cual utilizará el estudio de caso.**
   Diagnóstica, formativa o sumativa.
3) **Seleccione el estudio de caso.**
   Responda las siguientes preguntas: ¿el estudio de caso…
   - es congruente con el aprendizaje que pretendo evaluar?
   - es lo más cercano a la realidad?
   - es atractivo e implica un reto intelectual para su resolución?
   - se puede abordar en el tiempo del que se dispone?
   - brinda la oportunidad de recurrir a los conocimientos y la experiencia de mis alumnos?

4) **Defina quién realizará la evaluación.**
   Heteroevaluación, coevaluación o autoevaluación.
5) **Analice el texto, tomando en consideración las características de sus alumnos.**
   Adopte la postura de alumno y analice si la narración del estudio de caso es comprensible, si el caso es de su interés, qué elementos que se mencionan en él podrían desconocer, qué le dejaría el análisis del caso.
6) **Defina los criterios para evaluar el desempeño de sus alumnos.**
   Delimite los elementos que para usted son fundamentales valorar en el análisis del caso de acuerdo con el resultado de aprendizaje que desea evaluar.
7) **Seleccione el instrumento con el que evaluará el aprendizaje.**
   Dependiendo de la información que recabe, determine el instrumento que le será de

mayor utilidad para evaluar el aprendizaje de sus alumnos, por ejemplo, la rúbrica o la lista de cotejo.

Diseño

1) **Decida si adecua o no el caso original.**
En caso de considerarlo pertinente y más comprensible para sus alumnos, adecue el lenguaje del estudio de caso verificando el uso apropiado del lenguaje técnico y comprobando que tenga la información necesaria para realizar el análisis.

Consejo: Puede recurrir a casos polémicos que generen el involucramiento de los alumnos.

2) **Elija el formato en el que presentará el caso.**
Mediante hojas impresas o una presentación PowerPoint. Se sugiere proporcionar a los alumnos una copia para que puedan analizar el caso en el tiempo que ellos dispongan, así como investigar más a fondo, según lo consideren pertinente.
Algunos autores sugieren, de acuerdo con la Penn State University que los casos deben contener tres elementos: 1) un escenario real, 2) documentos y datos de soporte y 3) un problema de final abierto.

Sugerencia: El formato puede hacerse como postales, tarjetas, marcadores de páginas, etcétera.

3) **Elabore una tabla para poner sus anotaciones durante las dinámicas de trabajo.**
Como documento de trabajo, anote en una tabla el tipo de interacciones que surgieron en los grupos de trabajo. Esto le ayudará a obtener información valiosa sobre el trabajo en equipo: identificar a los alumnos que participan o son más retraídos, detectar conceptualizaciones incorrectas, etcétera.
4) **Entregue a sus alumnos los instrumentos con los criterios de evaluación.**
Entregue a sus alumnos el formato de los criterios con los que evaluará la interacción durante el análisis de caso.
5) **Determine si sus alumnos deben presentar un trabajo escrito.**
Revise la lista y haga los ajustes para su edición e impresión, ya sea en un formato de papel o en versión electrónica.

Revisión

- Es muy importante que durante el análisis del caso usted interactúe con cada uno de los grupos de trabajo, y así se percate de la dinámica existente entre los integrantes de los equipos.

- Durante el análisis del caso y después de la entrega del trabajo final, proporcione retroalimentación a los alumnos.
- Dialogue con los alumnos para conocer su opinión sobre este instrumento.

¿CÓMO LO APLICO?

Una vez que el profesor o facilitador identificó que el estudio de caso es la herramienta más apropiada para valorar el aprendizaje de los alumnos debe:

1) Proporcionar una breve introducción sobre esta herramienta (qué es, qué elementos que lo integran y cómo será evaluado).
2) Formar grupos de trabajo.
3) Presentar el caso a los grupos de trabajo.
4) Generar una lluvia de ideas en la que los alumnos expongan su opinión, sus juicios y las posibles soluciones para la problemática abordada.
5) Plantear una serie de preguntas que guíen el análisis del caso, por ejemplo:
    - ¿Cuál es el tema?
    - ¿Cuál es el contexto en el que se desarrolla la problemática?
    - ¿Cuáles son los factores clave que se deben considerar?
    - ¿Qué alternativas están disponibles para la toma de decisiones?
    - ¿Qué recomendaciones haría? ¿Por qué?
6) Promover la discusión entre los miembros del grupo de trabajo.
7) Solicitar a un representante de cada equipo que responda una de las preguntas planteadas en voz alta. Dependiendo del caso pueden plantearse en orden o aleatoriamente.

El estudio de caso también se puede analizar de forma individual, en tal caso el profesor o facilitador debe brindar seguimiento, asesoría y retroalimentación a los alumnos. Así, independientemente de que se analice grupal o individualmente, se deben plantear una serie de preguntas críticas que favorezcan la reflexión del alumnado, así como promover el intercambio de ideas continuamente.

El estudio de caso en la evaluación a distancia

Ante la situación actual de la pandemia por el COVID-19, se han dado avances notables en la implementación de cursos a distancia que anteriormente se llevaban de forma presencial, por lo que resulta indispensable que tanto docentes como estudiantes adquieran cada vez más herramientas para lograr el aprendizaje lo que permitirá que la evaluación se realice desde un enfoque formativo, además del sumativo, que es más común en los espacios educativos.

Como parte de estas necesidades de capacitación, el estudio de caso a distancia se puede convertir en un instrumento de evaluación muy útil en diversas áreas del conocimiento, ya que permite describir, analizar, comparar, evaluar y promover la reflexión de los estudiantes

sin necesidad de estar de forma presencial. Para esto se requiere seguir ciertos pasos que permita emplear esta herramienta de forma más eficiente.

¿Cómo lo diseño a distancia?

A diferencia de otras estrategias, el diseño del estudio de caso es prácticamente el mismo que en su aplicación presencial, solo se realizan los siguientes ajustes:

Planeación

- Identifique el o los objetivos de aprendizaje que quiere valorar
- ¿El estudio de caso es el mejor instrumento para valorar ese aprendizaje?
- Defina el tipo de evaluación para la cual utilizará el estudio de caso
- Diagnóstica, formativa o sumativa.
- Seleccione el estudio de caso

Responda las siguientes preguntas: ¿el estudio de caso…

  - es congruente con el aprendizaje que pretendo evaluar?
  - es lo más cercano a la realidad?
  - es atractivo e implica un reto intelectual para su resolución?
  - se puede abordar en el tiempo y con las herramientas de las que se dispone?
  - brinda la oportunidad de recurrir a los conocimientos y la experiencia de mis alumnos?

- Defina quién realizará la evaluación
- Heteroevaluación, coevaluación o autoevaluación.
- Analice el caso, tomando en consideración las características de sus alumnos
  Adopte la postura de alumno y analice si la narración del estudio de caso es comprensible, si el caso es de su interés, qué elementos que se mencionan en él podrían desconocer, qué le dejaría el análisis del caso. Puede complementar la información con imágenes, audios o videos para facilitar su comprensión.
- Defina los criterios para evaluar el desempeño de sus alumnos
  Delimite los elementos que para usted son fundamentales valorar en el análisis del caso de acuerdo con el resultado de aprendizaje que desea evaluar incluyendo el material audiovisual que acompañe el estudio.
- Seleccione el instrumento con el que evaluará el aprendizaje
  Dependiendo de la información que recabe, determine el instrumento que le será de mayor utilidad para evaluar el aprendizaje de sus alumnos, por ejemplo, la rúbrica y la lista de cotejo. Esta se puede compartir en línea con los estudiantes, generarse en un formulario Web o trabajarse como documento compartido en el ejercicio.

Diseño

- Decida si adecua o no el caso original
  En caso de considerarlo pertinente y más comprensible para sus alumnos, adecue el lenguaje del estudio de caso verificando el uso apropiado del lenguaje técnico y comprobando que tenga la información necesaria para realizar el análisis.
- Elija el formato en el que presentará el caso
  El caso puede ser proyectado a todos los estudiantes en formato de presentación, documento protegido, video, audio o un texto escaneado y puede complementarse incluso con otras fuentes como páginas web, por ejemplo. En algunos casos clásicos de psicología, por ejemplo, puede encontrarse información adicional en sitios web.

  Consejo: Puede recurrir a casos polémicos que generen el involucramiento de los alumnos.

- Elabore una tabla para poner sus anotaciones durante las dinámicas de trabajo.
  Como documento de trabajo, anote en una tabla el tipo de interacciones que surgieron en los grupos de trabajo como un documento compartido.
- Entregue a sus alumnos los instrumentos con los criterios de evaluación.
  Comparta con sus alumnos el documento con los criterios con los que evaluará la interacción durante el análisis de caso. Esto puede ser con un documento que se trabaje en línea en la sesión o incluso formularios web diseñados *exprofeso*.

  Sugerencia: El formato puede hacerse como postales, tarjetas, marcadores de páginas, etcétera.

Revisión

- Es muy importante que durante el análisis del caso usted interactúe con cada uno de los grupos de trabajo, y así se percate de la dinámica existente entre los integrantes de los equipos. En las aplicaciones como Zoom o Google Meet puede organizar grupos de trabajo e interactuar de forma individual con cada uno.
- Durante el análisis del caso y después de la entrega del trabajo final, proporcione retroalimentación a los alumnos.
- Dialogue con los alumnos para conocer su opinión sobre este instrumento.

¿Cómo lo instrumento a distancia?

El estudio de un caso en los entornos virtuales permite ampliar las estrategias de enseñanza-aprendizaje-evaluación en los espacios educativos, ya que puede promover la interacción e interconexión así como la creación de estudios de caso multimedia que lo pueden hacer más atractivo para los alumnos, sin perder las grandes ventajas que ofrece como el desarrollo de habilidades del pensamiento complejo relacionado con la reflexión que los alumnos deben hacer sobre diverso temas mediante un estudio de caso.

Una vez que el profesor o facilitador identificó que el estudio de caso es la herramienta más apropiada para valorar el aprendizaje de los alumnos debe:

1) **Proporcionar una breve introducción sobre esta herramienta (qué es, qué elementos lo integran y cómo será evaluado). Puede utilizar una presentación breve como guía.**
2) **Formar grupos de trabajo. En Zoom, Webex y Google Meet se pueden predeterminar grupos con los estudiantes si lo considera pertinente. No es recomendable hacerlo de forma asincrónica porque puede perderse el proceso de retroalimentación.**
3) **Presentar el caso a los grupos de trabajo.**
4) **Generar una lluvia de ideas en la que los alumnos expongan su opinión, sus juicios y las posibles soluciones para la problemática abordada. Puede hacerse en un chat o de manera verbal durante la sesión sincrónica.**
5) **Plantear una serie de preguntas que guíen el análisis del caso y proyectarlas en la sesión como documento compartido, por ejemplo:**
   a) ¿Cuál es el tema?
   b) ¿Cuál es el contexto en el que se desarrolla la problemática?
   c) ¿Cuáles son los factores clave que se deben considerar?
   d) ¿Qué alternativas están disponibles para la toma de decisiones?
   e) ¿Qué recomendaciones haría? ¿Por qué?

6) **Promover la discusión entre los miembros del grupo de trabajo.**
7) **Solicitar a un representante de cada equipo que responda una de las preguntas planteadas, de preferencia de forma verbal. Dependiendo del caso pueden plantearse en orden o aleatoriamente.**

Al igual que en la aplicación presencial, el estudio de caso también se puede analizar de forma individual, en tal situación, el profesor o facilitador debe brindar seguimiento, asesoría y retroalimentación a los alumnos. Así, independientemente de que se analice grupal o individualmente, se deben plantear una serie de preguntas críticas que favorezcan la reflexión del alumnado, así como promover el intercambio de ideas continuamente.

Ejercicio interactivo

En esta página se encuentra un estudio de caso que es útil como ejemplo para usarla con los estudiantes en clases o seminarios de motivación al estudio y fomentar su adhesión al estudio previo en contextos de aprendizaje inverso, lo invitamos a consultarla y crear un instrumento similar de acuerdo a su contexto como docente:

Estudio de caso Marta y José Luis para leer y reflexionar sobre las formas de abordar su aprendizaje (versión para estudiantes).

¿CÓMO LO ANALIZO?

La riqueza de esta herramienta se obtiene del análisis de las interacciones entre los integrantes de los grupos. La información que se genera durante la participación del alumnado puede valorarse cuantitativa y cualitativamente.

Para la evaluación cualitativa, se sugiere el empleo de instrumentos como la rúbrica y las listas de cotejo, los cuales permiten observar las conceptualizaciones, fortalezas y debilidades que tiene el alumno sobre un tema determinado. Estas pueden ser compartidas con los estudiantes para su conocimiento durante la sesión o pueden ser enviadas previamente por correo electrónico o mediante la plataforma utilizada para la enseñanza en línea.

Es fundamental que desde la planeación de la evaluación el profesor o facilitador, determine claramente los criterios a considerar en la revisión. Al igual que en la aplicación presencial, la evaluación se puede realizar de manera individual, por equipo o grupal.

Durante el proceso de análisis del estudio de caso, el profesor o facilitador debe resaltar las diversas formas de solución o enfrentamiento, al mismo tiempo que la importancia del proceso de razonamiento utilizado. Esto puede hacerse organizando las intervenciones de los estudiantes vía chat o de forma verbal por turnos con el micrófono abierto.

EJEMPLOS

Tabla 1. Lista de cotejo para evaluar la calidad del estudio de caso

Instrumento para el profesor o facilitador

Instrucciones: La siguiente lista de cotejo presenta algunos criterios que debe tener en cuenta en la selección o elaboración del estudio de caso que pretenda emplear con los alumnos. Si considera necesario anotar alguna observación importante, puede hacerlo en esa sección.

El estudio de caso que seleccioné o elaboré:	Sí	No	Observaciones
Es idóneo para evaluar el aprendizaje esperado.			
Considera las características de los alumnos.			
Despierta el interés de los alumnos.			
Se puede analizar en el tiempo que pretendo destinar para la actividad.			
Describe una situación concreta, basada en la realidad.			
Contiene toda la información prioritaria para su análisis.			
Describe de manera clara la metodología que se empleó.			
Recoge datos de distintas fuentes de información, que incrementan su validez.			
Incorpora las distintas perspectivas de las personas implicadas.			
Garantiza la confidencialidad de las personas o entidades implicadas.			
Favorece la adopción de una postura respecto al fenómeno de estudio.			
Promueve la toma de decisiones.			
Permite obtener conclusiones.			

Fuente: Elaboración propia.

Tabla 2. Rúbrica para evaluar el estudio de caso

Instrumento para el profesor o facilitador

Instrucciones: En la siguiente tabla se presentan tres niveles de desempeño de los alumnos respecto a tres etapas en el análisis de un estudio de caso: acercamiento, análisis y reflexión. Identifique el nivel que describe mejor el comportamiento de su alumno en cada una de las etapas. Si requiere precisar alguna información puede hacerlo en el apartado "Observaciones".

	Bueno	Regular	Malo
Acercamiento inicial al caso	El alumno recurre a gran cantidad de sus conocimientos previos para llegar a una solución. El alumno muestra mucho interés por el caso.	El alumno contempla algunos de sus conocimientos previos, pero muestra algunas dificultades para llegar a una solución. El alumno se muestra poco interesado por el caso.	El alumno no contempla ninguno de sus conocimientos previos, lo que dificulta mucho que llegue a una solución. El alumno no muestra interés por el caso.
Análisis del caso	El alumno plantea varias preguntas que guían su análisis del caso. El alumno resuelve las preguntas que plantea como resultado de la revisión de la literatura y las particularidades del caso. El alumno puede adoptar diferentes posturas y con base en ellas determinar cómo actuaría en esa situación. El alumno elabora conclusiones en las que incorpora posturas distintas a las que adoptó inicialmente.	El alumno plantea una pregunta para guiar el análisis del caso. El alumno responde la pregunta que plantea como resultado de la revisión de la literatura. El alumno contempla diferentes posturas, pero no determina la manera en que actuaría en esa situación. El alumno elabora conclusiones, pero en ellas no incorpora posturas distintas a las que adoptó.	El alumno no plantea ninguna pregunta que guíe el análisis del caso. El alumno no realiza ninguna revisión de la literatura. El alumno no adopta ninguna postura ni determina la manera en que actuaría en esa situación. El alumno no elabora conclusiones.
Reflexión final del caso	El alumno es capaz de reconocer las áreas en las que puede mejorar en el análisis del caso. El alumno es capaz de identificar de forma adecuada las estrategias que empleó en el análisis del caso.	El alumno reconoce algunas áreas en las que puede mejorar en el análisis del caso. El alumno identifica algunas estrategias que empleó en el análisis del caso.	El alumno no es capaz de reconocer las áreas en las que puede mejorar en el análisis del caso, ni las estrategias usadas para resolverlo.
Observaciones			

Fuente: Elaboración propia.

Tabla 3. Autoevaluación del análisis del estudio de caso

Instrucciones: A continuación, se presenta una serie de actividades que pudiste haber realizado durante y después del análisis del estudio de caso. Lee con atención cada una de ellas y marca con una cruz la casilla que mejor describa la frecuencia con que realizaste la actividad.

		Siempre	Casi siempre	Casi nunca	Nunca
Durante el análisis del caso	Identifiqué las características del caso.				
	Pude brindar una fundamentación teórica del caso.				
	Pregunté de manera clara mis dudas al profesor o facilitador.				
	Utilicé bibliografía para ahondar en el tema.				
	Elaboré interpretaciones sobre el caso.				
	Logré una comprensión profunda sobre el caso.				
Después del análisis del caso	Compartí mi opinión y mis reflexiones ante el grupo.				
	Elaboré conclusiones con los integrantes de mi grupo, que sintetizan los aspectos más relevantes del caso.				

Fuente: Elaboración propia.

RECURSOS EN LÍNEA

A continuación, se presentan algunas sugerencias de recursos en línea para ampliar la información de este capítulo:

¿Qué es un ESTUDIO DE CASO y cuáles son sus características? Cómo hacerlo y ejemplos.
Descripción: En este video se explican las características del estudio de caso, así como su aplicación en diferentes áreas con ejemplos prácticos que pueden ayudar a comprender como incluirlo dentro del proceso enseñanza-aprendizaje-evaluación.

Ejemplos de estudios de caso
Descripción: Página de la Universidad de Liverpool en la que se presentan interesantes estudios de caso.

Video "Estudio de caso" MOOC UNAM. En este video se explica en que consiste la Simulación y se describen sus principales caractéristicas: https://www.coursera.org/lecture/evaluación-cualitativa/estudio-de-caso-v5yu7

Estudio de Caso en Educación
Descripción: Página electrónica que provee información útil sobre las características del estudio de caso y sus posibles aplicaciones en el ámbito educativo.

VENNGAGE
Descripción: Se trata de una plataforma que facilita el diseño y elaboración de un estudio de caso de forma sencilla y visualmente más atractiva

CONCLUSIONES Y REFLEXIONES FINALES

- El estudio de caso como instrumento de evaluación debe contener información clara, descriptiva y suficiente que evidencie los conocimientos y habilidades que el estudiante emplea para su análisis y resolución.
- El planteamiento de preguntas críticas por parte del profesor o facilitador, el análisis riguroso, la autorreflexión y la exposición de opiniones de los alumnos deben acompañar siempre el estudio de caso.
- El empleo de este instrumento promueve el trabajo en equipo, el pensamiento crítico, la capacidad de síntesis, la resolución de problemas y la toma de decisiones.
- Es importante administrar adecuadamente el tiempo destinado al análisis del estudio de caso y acompañarlo de retroalimentación.

REFERENCIAS

Artile-Monteagudo, M., Artiles-Rivero, C. y Rodríguez-Gómez F. (2016). El estudio de casos como método problémico en ciencias médicas: una experiencia necesaria. *Edumecentro*, *8*(1), 165-173. Recuperado de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2077-28742016000100013&lng=es&tlng=es

Donoso-Vázquez, T. y Sánchez-Martí, A. (2014). *Orientación educativa y profesional. Estudio de casos*. Barcelona, España: ISEP. Recuperado de http://www.academia.edu/4400658/Orientaci%C3%B3n_educativa_y_profesional._Estudio_de_casos

Flores Hernández A y Veleros Valverde MC. (2010). Estudio de caso en entornos visuales como estrategia de enseñanza de los valores en la universidad. En Congreso Euro-Iberoamericano de Alfabetización Mediática y Culturas Digitales Sevilla: Universidad de Sevilla. (Internet) [Consultado el 27 de septiembre de 2021]. Recuperado de https://idus.us.es/handle/11441/56790

Martínez-Bonafé, J. (1988). El estudio de casos en la investigación educativa. *Investigación en la escuela*, 6, 41-50.

Martínez-Carazo, P. (2006). El método de estudio de caso: estrategia metodológica de la investigación científica. *Pensamiento & Gestión*, 20, 165-193. Recuperado de http://www.redalyc.org/articulo.oa?id=64602005

Merriam, S. (2009). Qualitative research: A guide to design an implementation. San Francisco, Estados Unidos: Jossey-Bass.

Stake, R. (2007). Investigación con estudio de casos. Madrid, España: Morata.

Center for teaching and learning. (2017). Using case studies to teach. Recuperado de http://www.bu.edu/ctl/teaching-resources/using-case-studies-to-teach/

UK Centre for Materials Education. (2017). *Teaching materials using case studies.* Recuperado de http://www.materials.ac.uk/guides/casestudies.asp

Capítulo 29

SOLUCIÓN DE PROBLEMAS

Ana Itzel Pascual Vigil, Cinthya Trejo Rojas,
Elibidú Ortega Sánchez, Adrián Martínez González

*"Un gran descubrimiento resuelve un gran problema,
pero hay una pizca de descubrimiento en la solución de cualquier problema."*
George Pólya

INTRODUCCIÓN

En la actualidad, la solución de problemas cobra un papel muy importante en la búsqueda de respuestas para distintas necesidades. La capacidad para dar solución a un problema resulta fundamental en todas las disciplinas, debido a que todo campo de conocimiento tiene situaciones que presentan dificultades para las cuales no hay soluciones evidentes (Krulik y Rudnik, 1980, citado por Becerra, Gras-Martí y Martínez-Torregrosa, 2004).

Los conocimientos que se generan en las disciplinas son extensos y profundos de tal manera que superan la capacidad de retención de cualquier individuo. Muchos de estos conocimientos en ocasiones no son suficientes para solucionar problemas complejos que ocurren en nuestro ambiente con gran incertidumbre. Por tanto, la tendencia en la educación se orienta al desarrollo de habilidades para identificar información idónea basada en la evidencia científica y en la utilización de esa información para solucionar problemas.

En el campo educativo, la solución de problemas permite que los alumnos construyan su aprendizaje, de forma presencial, mixta o a distancia, tomando como base tanto los problemas simulados como los de la vida real. Además, permite conocer más allá de los resultados de aprendizaje de los alumnos, ya que ayuda al profesor-facilitador a evaluar el razonamiento, el desarrollo y los avances que individualmente alcanzan los alumnos durante la formación.

La solución de problemas permite optimizar la evaluación de las habilidades que tiene el alumno para aplicar sus saberes, desarrollar estrategias y proponer una o varias soluciones a un problema. Además, brinda al profesor-facilitador un mayor conocimiento sobre las limitaciones, las necesidades y las potencialidades que poseen los alumnos de forma individual, grupal y generacional, a fin de conformar nuevas oportunidades de mejora del proceso de

enseñanza-aprendizaje. El proceso de solucionar problemas implica una serie de capacidades y habilidades del pensamiento que es importante desarrollar y evaluar en la preparación académica (Martínez-González, et al., 2006).

El propósito de este capítulo es orientar a los profesores-facilitadores en el uso de la solución de problemas como un recurso que los apoye en el diseño e implementación de la evaluación del, para y como aprendizaje cara a cara o en línea.

El capítulo está organizado en ocho secciones. En la primera se explica qué es la solución de problemas, en la segunda se detallan los pasos de su planeación y diseño, en la tercera se recomienda cómo aplicarla, en la cuarta se presenta cómo implementarla a distancia, en la quinta se señala cómo analizar sus resultados, en la sexta se incluye varios ejemplos para ilustrar su uso, en la séptima se presentan los recursos en línea, y en la octava se presentan algunas conclusiones y reflexiones finales.

¿QUÉ ES?

- La solución de problemas es un recurso de evaluación que busca identificar y comprender el proceso de razonamiento del alumno ante una situación particular, así como examinar las características de una propuesta de solución al problema planteado.
- Consiste en proporcionar una respuesta-producto a partir de un objeto o de una situación o en hallar una respuesta adecuada a las exigencias planteadas, sin embargo, la respuesta no debe verse como un logro final, sino como resultado de un proceso complejo de búsqueda, de encuentros, avances y retrocesos en el proceso de razonamiento (Cortés-Méndez y Galindo-Patiño, 2007).
- El problema es un recurso diseñado con el propósito de estimular el desarrollo de las habilidades del pensamiento para la solución de problemas y el proceso de enseñanza-aprendizaje, permitiendo que la información que brinde pueda ser considerada para realizar una evaluación diagnóstica, formativa o sumativa.
- Un problema puede plantearse mediante una pregunta, indicación o contradicción que no es posible contestar por la aplicación del sentido común, sino que, para solucionarse, es preciso poner en juego sus habilidades, activar el conocimiento previo, y establecer relaciones nuevas entre ellas.
- Los problemas planteados pueden hacer uso de diferentes fuentes para intentar motivar y despertar la curiosidad del alumno o de un grupo (Cortés-Méndez y Galindo-Patiño, 2007). Por ejemplo, puede utilizar recursos diferentes como casos reales, noticias nacionales e internacionales, series televisivas de actualidad, escenarios y tendencias sociales, entre otros.
- Lo más importante es que el problema sea visto como una herramienta para identificar la aplicación de las habilidades y los saberes, para detectar las necesidades y fortalezas tanto del estudiante, como del proceso de enseñanza-aprendizaje.
- Los problemas deben estar relacionados con una situación familiar ya resuelta o antes vista para que los conocimientos del alumno se activen y se proyecten sobre el nuevo

problema a solucionar (Coll, 1997, citado por Gorbaneff, 2010). Los problemas que diseñe el profesor-facilitador pueden ser estructurados o no estructurados.

- La solución de problemas posiciona al alumno frente a cinco grandes desafíos: a) la identificación de información relevante (datos y hechos orientadores), b) el planteamiento y la comprensión de un problema, c) la proposición de hipótesis o respuesta(s) al problema planteado, d) la determinación de estrategia(s) de solución o intervención y e) la aplicación de una solución o mejora del problema.

Figura 1



- El profesor-facilitador tiene la función de crear y recrear tanto situaciones como actividades pensadas para promover el uso de determinados contenidos, saberes, habilidades y actitudes por parte del alumno a partir de un problema que se puede desarrollar a corto, mediano o largo plazo. A continuación, explicaremos las consideraciones a fin de implementar la solución de problemas en una evaluación del, para y como aprendizaje de manera presencial, mixta y a distancia.

¿CÓMO LA DISEÑO?

Para implementar la solución de problemas se recomienda analizar el programa de su curso, módulo o asignatura para establecer los elementos que evaluará.

Después se sugiere seguir el siguiente proceso:

Planeación

1) **Determine los aprendizajes esperados que evaluará.**
   Considere objetivos, contenidos, actividades y aprendizajes esperados, así como las características de sus alumnos y el medio en el que se desarrollará el problema: físico-presencial, tecnológico a distancia con modalidad sincrónica o asincrónica.
2) **Seleccione el tipo de evaluación que desea realizar con la solución de problemas.**
   Puede ser diagnóstica si se quiere saber los conocimientos con los que cuenta el estudiante al inicio del curso, formativa si se quiere hacer énfasis en el aprendizaje y el proceso o sumativa si lo que se busca es valorar el resultado.
3) **Establezca quiénes participarán en la evaluación y de qué forma.**
   Puede elegir entre tres formas para realizar la evaluación: la primera en la que el agente de evaluación valora directamente la solución del problema (heteroevaluación); la segunda, donde un alumno o grupo de alumnos valoran el trabajo de otro alumno o grupo (coevaluación); y la tercera, en la cual el propio alumno con ayuda del docente valore su propio trabajo (autoevaluación).

   Consejo 1. Puede implementar una heteroevaluación, coevaluación, autoevaluación o una combinación de las anteriores.

4) **Seleccione el tipo de problema que se ajuste al objetivo de evaluación.**
   El problema estructurado presenta todos los aspectos de la situación, requiere la aplicación de reglas y posee una solución conocida. En contraste, los no estructurados se conforman de elementos que no son totalmente conocidos por el alumno, tienen diferentes soluciones y exigen la toma de decisiones y juicios de valor. En el caso de alumnos con más experiencia en solución de problemas, para valorar su capacidad de análisis en la identificación de errores y en la creación de estrategias para lidiar con ellos existe la posibilidad de proponer un problema de inicio mal estructurado.

Consejo 2. El escenario tiene que permitirle al alumno formularse dudas y realizar preguntas que lo guíen a la solución.

5) **Construya o seleccione con el problema un escenario o caso familiar.**
   El problema que plantee debe cubrir objetivos, contenidos, actividades y aprendizajes esperados.

   Consejo 3. Considere si es pertinente el uso de temas actuales, controvertidos y relacionados con la edad de los alumnos.

6) **Determine la solución esperada al problema planteado.**
   La solución esperada al problema considerará el proceso que debe realizar el alumno con las características y contenidos vistos en el curso, módulo o asignatura.
7) **Seleccione un instrumento que acompañe la evaluación del proceso de solución de problemas.**
   El instrumento puede ser el modelo de respuesta correcta, rúbrica, lista de cotejo, escala u otro instrumento diseñado a partir de los criterios de evaluación que determine.

Diseño

Para la presentación visual de la solución de problemas, se sugiere utilizar tres formatos: uno para el alumno, otro para el profesor-facilitador y un instrumento de apoyo para la evaluación.

Formato para el alumno (Ver Ejemplo 1)

1) **Defina el título del problema.**
   Considere un título llamativo para atrapar la atención del alumno: ¿Eres lo que comes?
2) **Incluya datos de identificación.**
   Unidad temática a la que corresponde el problema, datos del alumno y fecha.
3) **Precise las instrucciones para orientar al alumno en la actividad.**
   Explique de manera sistemática las actividades que realizará el alumno, lo que espera como producto final y el tiempo que tiene para responderlo.
4) **Presente el problema que el alumno debe aclarar o solucionar.**
   Los alumnos según su nivel de conocimientos pueden al inicio solamente aclarar el problema planteado y conforme van aumentando sus conocimientos pueden llegar poco a poco a su solución. Verifique que el caso o escenario del problema contiene toda la información necesaria para que el alumno pueda plantear una respuesta.

   Consejo 4. El escenario o caso del problema debe llevar al alumno a tomar decisiones o hacer juicios basados en hechos, información lógica y fundamentada.

Formato para el profesor-facilitador (Ver Ejemplo 2)

Incluya los pasos del 1 al 4 e incorpore los que se encuentran a continuación (Ver Ejemplo 2)

5) **Establezca los datos y hechos orientadores para identificar y plantear lo(s) problema(s).** Seleccione los hechos, datos, condiciones o aspectos indispensables que caracterizan al escenario o caso con la finalidad de identificar el problema(s).
6) **Plantee los problemas en forma de preguntas con base en los hechos y datos.** Son las preguntas que deben identificar y plantear los alumnos a partir del escenario (hechos y datos) que se les presentó.
7) **Formule hipótesis o explicaciones que den respuesta a la pregunta(s) planteada(s).** Son posibles explicaciones que el estudiante tendrá que fundamentar cuando vaya a consultar las diversas fuentes de información.
8) **Incorpore las estrategias que se evaluarán.** Estos planteamientos deberán identificarlos los estudiantes con apoyo del profesor-facilitador para saber qué y dónde buscar y así fundamentar las hipótesis o respuestas planteadas a las preguntas.
9) **Fuentes de información.** Oriente a los alumnos de las diversas fuentes de información que puede utilizar para fundamentar sus respuestas.

   Consejo 5. Puede incluir las fuentes de información que espera retomen sus alumnos o preguntas de estímulo para apoyarlos.

10) **Preguntas de estímulo.** Se plantean con el fin de apoyar a los profesores-facilitadores para utilizarlas en el momento oportuno del proceso de análisis del problema.
11) **Defina una solución posible al problema o escenario presentado.** Esta solución dependerá del tipo de problema que seleccionó. Por ejemplo, si fue un problema estructurado con una única solución, debe trazar el proceso que el alumno debió seguir para llegar a la solución.

    Consejo 6. Se recomienda realizar el ejercicio interactivo para dar un repaso a las fases de la solución de problemas.

Formato de acompañamiento para la evaluación (Ver Ejemplos 3, 4 y 5)

1) **Seleccione el instrumento que acompañará a la evaluación de la solución del problema.** Puede utilizar una rúbrica, lista de cotejo, escala u otro instrumento.

   Consejo 7. Se recomienda que el instrumento que utilice incluya un apartado de observaciones para que pueda registrar elementos fundamentales del proceso que desarrolló el alumno para solucionar el problema.

2) **Establezca los criterios de evaluación del problema.**
   Puede considerar como criterios la comprensión del problema, la claridad en las propuestas de solución y el uso de recursos para recopilar información, entre otros.
3) **Diseñe su instrumento de evaluación.**
   Para diseñarlos puede consultar los capítulos sobre estos instrumentos: rúbrica, lista de cotejo u otros.

   Consejo 8. Muchos problemas se pueden solucionar de distintas formas y de acuerdo con distintos parámetros. Usted tendrá que valorar las propuestas de cada alumno con una serie de criterios.

4) **Verifique su instrumento de evaluación.**
   Realice una revisión final al instrumento antes de usarlo para evitar faltas de ortografía o errores en la redacción.

Revisión

- Un ejercicio para enriquecer la solución de problemas es realizar los casos o escenarios con otros profesores-facilitadores, se sugiere que sean alrededor de cinco, a fin de recibir observaciones y sugerencias que permitan enriquecer este recurso.
- Después de la aplicación, es muy útil que pida a los alumnos comentarios o sugerencias sobre el caso o escenario, el tiempo de ejecución y sus aprendizajes alcanzados, entre otros aspectos para mejorar el instrumento.

¿CÓMO LA APLICO?

Para aplicar la solución de problemas en los espacios educativos se debe:

- Diseñar problemas que permitan cubrir los objetivos del curso, módulo o asignatura. Puede considerar que se resuelvan en diferentes espacios como: el aula, un laboratorio, una comunidad o en línea.
- Identificar los momentos más oportunos para aplicar los problemas de acuerdo con el propósito de la evaluación.
- Definir si la aplicación será individual o grupal.
- Considerar el tiempo que necesitan los alumnos para realizar el trabajo de aclaración o solución.
- Presentar el problema diseñado en el momento planeado con el formato y una explicación de lo que espera de los alumnos. Es importante que comparta con los alumnos el instrumento de evaluación para que conozcan cómo serán evaluados y orienten su proceso de aprendizaje.

Consejo 9. Los medios que puede usar para comunicar el problema dependerán de su espacio de aprendizaje. Puede ser impreso o en formato digital.

- Proporcionar preguntas adicionales relacionadas con el problema que orienten al alumno en la solución. Por ejemplo, cuando percibe que los alumnos tienen dificultades para sugerir estrategias de solución, puede incorporar preguntas guía que los orienten de forma efectiva a encontrar una posible respuesta.
- Facilitar el aprendizaje con las respuestas a las dudas que surjan de los alumnos durante el proceso educativo, se sorprenderá de lo que son capaces de aclarar y solucionar. Recuerde que una situación nueva puede requerir de acompañamiento.
- Sugerir nuevas dinámicas en su espacio educativo como:
    - Modificaciones a las soluciones propuestas por los alumnos con base en las observaciones realizadas.
    - Realizar combinaciones entre las soluciones propuestas individualmente, con el fin de establecer relaciones entre los razonamientos de los alumnos y enriquecer el aprendizaje individual y grupal.

LA SOLUCIÓN DE PROBLEMAS EN LA EVALUACIÓN A DISTANCIA

¿Cómo la diseño?

Para el diseño de la solución de problemas en la evaluación a distancia, la planeación mantiene un proceso similar, sin embargo, deberá considerar lo siguiente:

- Examine el medio tecnológico en el que se presentará el problema, dependerá del medio el tipo de formato para el alumno y su implementación.
- Verifique que el medio que seleccionó es compatible con su espacio de trabajo virtual y que domina su uso para atender cualquier problema que surja.
- Contemple el tiempo que le llevará subir el problema en el medio seleccionado.
- Establezca el tipo de comunicación que mantendrá con los alumnos durante la solución del problema (sincrónica o asincrónica). Puede considerar las siguientes preguntas: durante la solución del problema ¿es necesaria mi presencia para que mis alumnos puedan solucionarlo?, ¿qué tipo de información sobre el aprendizaje puedo obtener al estar presente cuando solucionen el problema?
- Una vez que establezca el tipo de comunicación considere crear reglas de interacción. Por ejemplo: si es un grupo de trabajo de forma asincrónica, todos deberán participar por lo menos una vez en un espacio establecido. O si se trata de un trabajo sincrónico puede determinar el orden de las participaciones y el número de consultas que puede hacerle el alumno.
- Contemple que no todos los alumnos desarrollarán el proceso esperado debido al acceso limitado a infraestructura, *software* y *hardware*. Bajo este escenario, considere actividades

que el alumno pueda desarrollar en su contexto para darle una oportunidad de aprendizaje y evitar su rezago.

- Fomente retroalimentaciones sincrónicas y asincrónicas con sus alumnos. Las primeras sirven para atender dudas emergentes y promueven la presencia social y la socialización del aprendizaje, mientras que las segundas proporcionan una guía que el estudiante puede utilizar en el momento en que emprende su proceso de aprendizaje de manera independiente.

¿Cómo la aplico?

- La solución de problemas puede emplearse en distintas plataformas o aplicaciones, pero es importante que la selección la realice con base en:

  - el tiempo del que disponen los estudiantes para la estrategia, por ejemplo, si optará por llevarla a cabo como una actividad síncrona o asíncrona.
  - la disposición de recursos para implementarla, es decir, si las herramientas requieren de conexión a internet o *software* especializado, opte por aquéllas que son de acceso libre.
  - el tipo de supervisión durante la realización de la estrategia, por ejemplo, si será en el tiempo efectivo de clase, como actividad complementaria o como tarea fuera del horario regular.

  Consejo 10. Por ejemplo, puede utilizar Genially para presentar el problema y solicitar las respuestas abiertas por medio de un formulario en Google Forms: https://view.genial.ly/612585cce622bb0d2e3625b8/interactive-image-imagen-interactiva

- Logre comunicar adecuadamente los pasos que seguirá el alumno para la solución de problemas en la herramienta que seleccione, centrándose en el sentido que cada problema tiene para la construcción del aprendizaje que se quiere lograr.
- Previo a la aplicación, es indispensable que realice una prueba de la herramienta que utilizará desde el rol del estudiante, para verificar la vista del usuario, la redacción y gramática del contenido, el tiempo del que se dispone y el correcto funcionamiento de la herramienta.
- Recuerde que el uso de cualquier herramienta requiere de una guía sobre su funcionamiento para los alumnos como usuarios, para que su desempeño no se vea afectado. Por ello, es importante que usted la conozca desde el rol del alumno.
- Existe una gran cantidad de recursos que puede consultar y adaptar para aplicar la solución de problemas, recuerde que algunas son gratuitas y esto permite potenciar su creatividad.

¿CÓMO ANALIZO SUS RESULTADOS?

- La solución de problemas permite obtener información sobre el proceso de razonamiento del alumno para solucionar problemas. Para que pueda evaluarla, se debe realizar una serie de acciones:

  - Determinar si el alumno identificó los hechos o datos que orientaron el problema.
  - Establecer si el alumno identificó el problema. Planteó el problema en forma de pregunta.
  - Valorar si el alumno formuló hipótesis o respuestas posibles.
  - Identificar si determinó las estrategias esperadas para solucionar el problema.
  - Valorar si las fuentes consultadas por el alumno fueron las adecuadas.
  - Definir si fue exitosa la solución del problema.

- Para llevar a cabo este análisis se recomienda hacer uso de un registro, en el cual el profesor-facilitador realice observaciones sobre el proceso que desarrolló el alumno para solucionar el problema, centrando su atención en los logros y en las dificultades que presentó. Algunas preguntas que pueden orientar esta tarea son: cómo lo solucionó, qué utilizó para solucionarlo y cuál fue el proceso para llegar a ese resultado.
- Una vez que se determina el desempeño del alumno utilizando el instrumento de acompañamiento para la evaluación, se pueden incluir sugerencias sobre el proceso que siguió el estudiante para responder el problema. Estas sugerencias se las compartirá al alumno para tener una retroalimentación sobre las dificultades que se observaron en su proceso de solución, con el fin de estimularlo y lograr un nivel más profundo en la comprensión de los problemas abordados.
- El análisis puede ser realizado por el profesor-facilitador, un grupo de profesores, por el alumno o por un grupo de alumnos, según el propósito de su evaluación. Por ejemplo, para un propósito formativo, se puede utilizar la coevaluación en su grupo para que los alumnos valoren e identifiquen las diferentes propuestas de sus compañeros a un mismo problema, lo cual permite enriquecer las perspectivas sobre las que se puede abordar un problema.
- A nivel individual o en equipos de trabajo, se puede valorar el razonamiento de cada uno de los alumnos.
- De manera grupal, se puede proponer compartir las soluciones más eficientes que encontraron algunos alumnos o equipos de trabajo. De esta forma, se puede hacer evidente que existen diferentes formas de solucionar un mismo problema.

Si la aplicación de la solución de problemas se llevó a cabo en línea el profesor-facilitador debe tomar en cuenta, además de lo que se incluye en el apartado anterior, lo siguiente:

- Elija e indique a los alumnos el tipo de comunicación que mantendrá (sincrónica o asincrónica) para hacer la retroalimentación sobre sus resultados.

- La retroalimentación en línea individual o grupal debe ser más concisa: recuerde enfocarse en cómo resolvió el problema, qué empleó para solucionarlo y cuál fue el proceso para llegar a ese resultado para ofrecerle opciones de mejora en su comprensión, método y solución del problema.
- En la retroalimentación de este tipo de instrumentos es posible ir más allá de los resultados de aprendizaje esperados, por lo que debe dirigirse más a la capacidad de razonamiento de los alumnos, en el desarrollo y los avances que han logrado durante su proceso de formación.
- Al realizar retroalimentaciones recuerde que el medio donde se desarrolló la aplicación puede interferir en los resultados del alumno. Por ejemplo, si los alumnos no están familiarizados con la plataforma en donde van a solucionar el problema puede que su desempeño se vea empobrecido y se pierda seguridad sobre el grado real de avance en su aprendizaje.
- Cuando no todos los alumnos tienen un acceso regular a la infraestructura, *software* y *hardware*, procure complementar la evaluación con otros instrumentos.

  Consejo 11. En aplicaciones a distancia considere que éste no debe ser el único instrumento de evaluación.

- El instrumento específico que se empleó para complementar la evaluación ayuda a lograr objetividad y precisión, por lo que es necesario retomarlo para garantizar una retroalimentación efectiva.

EJEMPLOS

Ejemplo 1
Formato para el alumno

¿Eres lo que comes?

Problema de la Unidad 1: La prevención en el autocuidado de la salud
Nombre del alumno: ______________________________ Fecha: ____________________

Instrucciones: Lee con atención el siguiente párrafo y determina cuál es el problema planteado, cuentas con dos días para entregar tu respuesta.

Yo soy un(a) estudiante de bachillerato y me preocupa saber si mi alimentación reúne las características de una dieta ideal. Tengo la impresión de que lo que ingiero no es lo más adecuado para mis requerimientos energéticos, de acuerdo con mi:
Sexo: ____
Estatura: ____
Edad: ____
Peso actual: ____

Peso ideal: ____

Actividad que desarrollo: ____

¿Será que realmente estoy siguiendo una dieta ideal?

Si no estoy realizando una dieta ideal, ¿cuáles serían las características para poder iniciar una dieta ideal?

Mis datos y hechos orientadores para identificar el problema son:

____

El problema(s) que identifico es:

____

Mi hipótesis o explicaciones son:

____

Mis estrategias para solucionar el problema son:

____

La solución al problema que identifiqué es:

____

Las fuentes de información que utilicé fueron:

____

Ejemplo 2
Formato para profesores-facilitadores

¿Eres lo que comes?

Problema de la Unidad 1: La prevención en el autocuidado de la salud

Nombre del alumno: ________________________________ Fecha: _____________

Instrucciones: Lee con atención el siguiente párrafo y determina cuál es el problema planteado, cuentas con dos días para entregar tu respuesta.

Yo soy un(a) estudiante de bachillerato y me preocupa saber si mi alimentación reúne las características de una dieta ideal; tengo la impresión de que lo que ingiero no es lo más adecuado para mis requerimientos energéticos, de acuerdo con mi:

Sexo: ____

Estatura:____

Edad:____

Peso actual:____

Peso ideal:____

Actividad que desarrollo:____

¿Será que realmente estoy siguiendo una dieta ideal?
Si no estoy realizando una dieta ideal, ¿cuáles serían las características para poder iniciar una dieta ideal?

Propósito del problema
El alumno evaluará su dieta habitual y elaborará una nueva con base en las características de la dieta ideal.
Hechos o datos orientadores
Mi alumno utilizará los siguientes datos para determinar su dieta ideal:
- Sexo
- Edad
- Peso actual
- Peso ideal
- Estatura
- Actividad que desarrolla
- Cálculo del índice de masa corporal

Problema(s)
Mi alumno da respuesta a las siguientes preguntas:
- ¿Su dieta cumple con las características de una dieta ideal?
- ¿Qué pasos debe seguir para hacer una evaluación de su dieta?

Hipótesis o explicaciones
Mi alumno puede solucionar las preguntas considerando:
- El cálculo del índice de masa corporal, de las calorías, proteínas, hidratos de carbono, lípidos, agua y fibra que ingiere.
- La comparación de su dieta habitual con la dieta ideal que debe tener.
- El registro durante tres días de su dieta habitual o el recordatorio de 24 horas.

Estrategias
Mi alumno para solucionar el problema puede:
- Elaborar un informe que contenga el concepto de dieta y las características de la dieta ideal.
- Calcular el requerimiento energético diario, así como de proteínas, hidratos de carbono, lípidos, agua y fibra que ingiere.
- Realizar la evaluación de su dieta y del índice de masa corporal.
- Elaborar el diagnóstico de su dieta.
- Diseñar las recomendaciones de modificación de su dieta actual, ajustándola a lo que establece la dieta ideal, de acuerdo con sus características antropométricas y actividad que desarrolla.
- Determinar qué alimentos debe evitar y cuáles incluir o incrementar.

Fuentes de información
Internet, artículos de investigación en Pubmed, libros de Medicina Interna, nutriólogos, nutrición humana, etc.

Preguntas de estímulo

¿Qué es alimentación?, ¿qué es alimento?, ¿qué es nutrición?, ¿qué es dieta?, ¿cómo se calculan los requerimientos energéticos de un adulto?, ¿cómo se calcula el índice de masa corporal en un adulto?

Respuesta al problema

Mi alumno debe:

- Evaluar si su dieta reúne las características de una dieta recomendable (o si no las reúne).
- Elaborar su dieta ajustándola, de acuerdo con sus características antropométricas y actividad que desarrolla, a lo que establece una dieta ideal.

Ejemplo 3
Formato de acompañamiento para la evaluación

Lista de cotejo para la solución de problemas

Problema de la Unidad: ____

Nombre del evaluado: ________________________ Fecha: _______________

A continuación, se presentan seis criterios para la evaluación de la solución del problema planteado. Marque con una X si considera que la propuesta del alumno cumple o no con cada criterio y asigne un punto por cada criterio marcado en SÍ.

Evaluación para la solución de problemas	Sí	No
El alumno identificó los hechos y datos orientadores.		
El alumno identificó el problema o los problemas.		
El alumno formuló hipótesis o respuestas posibles.		
El alumno identificó las estrategias para fundamentar sus respuestas.		
El alumno utilizó fuentes de información adecuadas.		
La solución al problema con las estrategias empleadas por el alumno fue exitosa.		

Nombre del evaluador: _______________

Observaciones ____

____

____

Fuente: Elaboración propia.

Ejemplo 4
Formato de acompañamiento para la evaluación

Registro de observaciones para la solución de problemas

Problema de la Unidad:____
Nombre del evaluado: ____________________________ Fecha: ___________________

El siguiente registro de observaciones contiene cinco elementos guía, en los cuales puede describir las fortalezas y/o dificultades que percibió en la propuesta del alumno al problema planteado.

Elementos a observar	Retroalimentación	
	Fortalezas	Dificultades
Identificación de hechos o datos orientadores.		
Identificación de problema(s) en forma de pregunta.		
Planteamiento de hipótesis o respuestas a la pregunta planteada.		
Identificación de estrategias.		
Uso de las fuentes de información.		
Respuestas al problema.		

NOTA: El formato de registro de observaciones es un buen complemento para acompañar la evaluación.

Nombre del evaluador: ____

Fuente: Elaboración propia.

Ejemplo 5
Formato de acompañamiento para la evaluación

Autoevaluación para la solución de problemas

Instrucciones: Contesta las siguientes preguntas reflexionando sobre el análisis, desarrollo y solución del problema planteado.

1) ¿Por qué propusiste esa solución al problema planteado? ¿Qué criterio utilizaste?

____

____

____

2) Al realizar el análisis del problema, ¿qué fue lo más importante para ti?

____

____

____

3) ¿En qué es similar o diferente a otros problemas abordados en clase?

____

____

____

4) ¿Cuáles son, según tu criterio, las áreas fuertes y débiles de tu solución al problema?

____

____

____

5) Si pudieras volver a solucionar el problema, ¿qué harías?, ¿cambiarías algo?

____

____

____

Fuente: Elaboración propia.

RECURSOS EN LÍNEA

Nombre	Tipo	Descripción
Educaplay https://es.educaplay.com/	Plataforma	Es una herramienta gratuita en línea que permite utilizar o generar recursos y crear actividades como mapas interactivos, memoramas, crucigramas, sopa de letras, entre otros, es fácil de compartir si deseas conocer más: https://youtu.be/IJ2bZYMwZnM
Educlipper https://app.participate.com/	Plataforma	Es una herramienta gratuita que permite explorar y compartir recursos de contenido educativo, por ejemplo, videos, imágenes, material sonoro, etc.
George Pólya - Estrategias de Resolución de Problemas Fernando Faundez	Video	El video nos permite conocer la propuesta teórica de solución de problemas con base en el método desarrollado por George Pólya (Licencia Atribución de Creative Commons (reutilización permitida)) https://www.youtube.com/watch?v=odXMHOhcmQA
¿Cómo se resuelve un problema? Alejandro Candioti	Video	En el video nos narra de qué forma podemos solucionar problemas complejos por medio de analogías más sencillas, que pueden contribuir a desarrollar ideas que ayuden a resolver problemas más difíciles. https://youtu.be/mqWPS9LkFvo
Me gustan los problemas José Ángel Murcia	Video	En el video nos explican que no hay una respuesta estándar para los problemas y que, para solucionarlos, existe una variedad de soluciones que dependen de la creatividad y el pensamiento de las personas que lo resuelven. https://youtu.be/agiRtqE9WDk
¿Sabemos resolver problemas? Martín Pérez Mendóza	Video	El video nos invita a reflexionar sobre la causa de los problemas como estrategia para ver más allá de los síntomas de las problemáticas para atacar las causas de raíz. https://www.youtube.com/watch?v=_-VaX2y25gU
Ejercicio sobre solución de problemas	Ejercicio interactivo	Breve ejercicio para repasar las fases de la solución de problemas como herramienta de evaluación. https://youtu.be/Hr3NgtMr91c

Laberinto sobre la planeación de solución de problemas	Ejercicio interactivo	Ejercicio sobre la ruta que sigue el proceso de diseño para utilizar la solución de problemas en un proceso de evaluación. Introduce la contraseña teniendo en cuenta el número de flechas que has pasado para llegar hasta ahí. https://bit.ly/3NoxCUP

CONCLUSIONES Y REFLEXIONES FINALES

- La solución de problemas de forma presencial, mixta o a distancia, promueve procesos cognitivos complejos de alto nivel como el pensamiento crítico, reflexivo, el razonamiento y la argumentación utilizados para fundamentar la solución al problema.
- Impulsa la creatividad para diseñar soluciones debido a la libertad e interacción que tienen alumno-profesor-grupo.
- Contribuye a que el profesor-facilitador identifique los puntos débiles y fuertes de la aplicación del aprendizaje y retroalimentaciones al alumno para que ambos puedan aplicar medidas de mejora.
- Se sugiere una construcción cuidadosa e ingeniosa del problema, para que la tarea sea atractiva y desafiante para los alumnos.
- Se recomienda que los criterios a evaluar y el problema estén conjugados para lograr una evaluación efectiva.
- La solución de problemas permite la observación del desempeño del alumno en la producción de un producto complejo y posibilita realizar un juicio sobre su calidad.

REFERENCIAS

Arias, N. González, K. y Padilla, J. (2010). Educación a distancia y educación virtual: una diferencia necesaria desde la perspectiva pedagógica y la formación del ser humano. *Revista de Investigaciones UNAD*. *9*(3) Recuperado de https://hemeroteca.unad.edu.co/index.php/revista-de-investigaciones-unad/article/view/726

Becerra, C. Gras-Martí, A. y Martínez-Torregrosa, J. (2004). Análisis de la solución de problemas de física en secundaria y primer curso universitario en Chile. *Enseñanza de las ciencias*. *22*(2).275-286. Recuperado de https://www.raco.cat/index.php/Ensenanza/article/view/21978/21812

Bastién, G., Pérez, A. y Rodríguez A. (2016). Epistemología para la resolución de problemas a distancia. *Suplemento SIGNOS EAD*. Recuperado de https://p3.usal.edu.ar/index.php/supsignosead/article/view/3704/4596

Cortés-Méndez, M. y Galindo-Patiño, N. (2007). *El modelo de Pólya centrado en resolución de problemas en la interpretación y manejo de la integral definida. Un estudio realizado con estudiantes de Ingeniería del Grupo 07 de segundo semestre del año 2006 de la Universidad La*

Salle, Trabajo de grado para optar el título de Maestría en Docencia. Bogotá, Colombia: Universidad La Salle. Recuperado de https://bit.ly/38S9fjd

Font-Ribas, A. (2004). Líneas maestras del aprendizaje por problemas, *Interuniversitaria de Formación del Profesorado*, *18*(1). 79-95. Recuperado de http://www.redalyc.org/articulo.oa?id=27418106

García-García, J. y Rentería-Rodríguez, E. (2013). Resolver problemas: una estrategia para el aprendizaje de la termodinámica. *Revista Científica Guillermo de Ockham*, *11*(2), 117-134.

Gorbaneff, Y. (2010). Qué se puede aprender de la literatura sobre el aprendizaje basado en problemas, *Revista Facultad de Ciencias Económicas: Investigación y Reflexión*, *8*(1), 61-74. Recuperado de http://www.redalyc.org/articulo.oa?id=90920479003

Gutiérrez, A., Puente, G., Martínez, A. y Piña, E. (2012). Aprendizaje basado en problemas un camino para aprender a aprender. CCH-UNAM. 1ª ed. diciembre 2012. ISBN: 9786070239021

Koehler, A., Cheng, Z., Fiock, H., Janakiraman, S., y Wang, H. (2020). Asynchronous online discussions during case-based learning: A problem-solving process. *Online Learning, 24*(4), 64-92. Recuperado de https://eric.ed.gov/?id=EJ1277718

Martínez-González, A., Gutiérrez-Ávila H. y Piña Garza E. (2006) *Aprendizaje Basado en Problemas en la Enseñanza de la Medicina y Ciencias de la Salud*. México, Editorial: Facultad de Medicina, UNAM: 1ª ed. mayo de 2006. ISBN: 9703233058.

Nieto, S. y José H. (2005). Resolución de problemas, Matemáticas y Computación. *Revista Venezolana de Información, Tecnología y Conocimiento*, *2*(2), 37-45. Recuperado de http://www.redalyc.org/articulo.oa?id=82320204

Rahman, M. (2019). 21st Century Skill "Problem Solving": Defining the Concept. Asian Journal of Interdisciplinary Research, *2*(1), 71-81. Recuperado de https://eric.ed.gov/?id=ED593994

Sousa, F. Monteiro, I. y Pellissier, R. (2014). The development of a small-word network of Higher Education Students, using a large-group Problem-Solving Method. *Research on Quality of Higher Education*, 11, 68-87. Recuperado de https://eric.ed.gov/?id=EJ1065828

Capítulo 30

PROYECTO

María Juliana Londoño Cárdenas, María Elena Pérez Rivera, Adrián Martínez González

"El arte supremo del maestro es
despertar el placer de la expresión creativa y el conocimiento."
Albert Einstein

INTRODUCCIÓN

La evaluación basada en proyecto puede ser un proceso complejo, colaborativo y multidireccional, en el cual los estudiantes y profesores aprenden a trabajar de manera conjunta e individual para lograr objetivos específicos (Ahumada-Acevedo, 2005). Es una de las estrategias que se utiliza para que los estudiantes tengan la oportunidad de adquirir y desarrollar conocimientos y habilidades durante el periodo de tiempo que investiga y resuelve un problema, reto o pregunta compleja.

Para que la evaluación resulte pertinente es necesario alinearla con el mapa curricular, perfil de egreso, programa de asignatura y la enseñanza. Con esta estrategia –ya sea implementada en una modalidad presencial, a distancia o mixta– se puede valorar la resolución de problemas, el pensamiento crítico, la creatividad, las habilidades de comunicación, colaboración, autonomía, toma de decisiones, entre otros. Los estudiantes y docentes reflexionan sobre el aprendizaje, la efectividad de sus preguntas y actividades del proyecto, la calidad del trabajo, las dificultades que se encontraron y la forma en que se resolvieron (BIE, 2017).

Este capítulo presenta información útil para la implementación de una evaluación del y para el aprendizaje basada en proyectos. Se organiza en ocho secciones. En la primera se explica qué es una evaluación basada en proyectos, en la segunda se detallan los pasos para diseñarlo, en la tercera se recomienda cómo aplicarlo, en la cuarta se proponen elementos para su implementación en la modalidad a distancia, en la quinta se sugiere cómo analizar sus resultados, en la sexta se proponen tres ejemplos de instrumentos de recolección de información, en la séptima se incluyen algunos recursos en línea, y en la octava se presentan conclusiones y reflexiones finales.

¿QUÉ ES?

- Es una estrategia que pretende valorar progresivamente el desempeño de los estudiantes a partir de la asignación de un proyecto que implique el uso de habilidades cognitivas, personales para la resolución de problemas o la elaboración de un producto.
- Un proyecto busca enfrentar a los estudiantes a situaciones que los lleven a rescatar, comprender y aplicar aquello que aprenden como una herramienta para resolver problemas o proponer mejoras en las comunidades en donde se desenvuelven (Instituto Tecnológico y de Estudios Superiores de Monterrey, 2000).
- La evaluación basada en proyectos es versátil, flexible y adaptable a diferentes escenarios porque su planificación y desarrollo dependen de la creatividad de los estudiantes y del profesor.
- Se pueden considerar diferentes tipos de proyectos (Brown, Bull y Pendlebury, 1997):
    - **Experimentales**: son aquellos que responden una pregunta de investigación, emplean rigurosamente el método científico.
    - **Documentales**: responden una pregunta de investigación basándose en los aportes de la literatura sobre un tema relevante.
    - **De taller** *(studio-based)*: se desarrollan de forma individual en un espacio común y se expone el tema de forma pública.
    - **Laborales** *(work-based)*: se realizan en colaboración con organizaciones externas.
    - **Comunitarios**: dan énfasis a la responsabilidad social y a la conciencia cultural.
- La evaluación basada en proyectos se puede realizar a lo largo o al final del curso, y se puede proponer que el proyecto se desarrolle a nivel grupal o individual, pero será decisión del docente si la calificación será de manera individual o grupal.
- Es una estrategia que se emplea en todos los niveles educativos, ya que ayuda al estudiante a apropiarse de su aprendizaje. En el caso de los posgrados con orientación en investigación esta es la estrategia más utilizada.
- Forma parte del Aprendizaje Basado en Proyectos (ABP).[1]
- Uno de los propósitos de los proyectos es el desarrollo de habilidades de investigación en distintos contextos, académicos y laborales (Brown, et al., 1997).
- Una limitación de la elaboración de un proyecto es el tiempo en que se puede realizar, ya que por lo regular es restringido. En el caso de la evaluación de un proyecto, un obstáculo es la forma en que se dé la retroalimentación, puesto que en ocasiones es más subjetiva que objetiva.

[1] Canal Pontificia Universidad Católica del Perú - PUCP. (21 de noviembre de 2017). ¿Qué es el Aprendizaje Basado en Proyectos? [Archivo de Vídeo]. Youtube. https://www.youtube.com/watch?v=Vp4ZQu4T6lw

¿CÓMO LA DISEÑO?

Para realizar una evaluación basada en proyectos, es necesario analizar el mapa curricular, el perfil de egreso y el programa de asignatura para considerar qué proyecto se desea realizar. El desarrollo de este permite que tanto el conocimiento aprendido como su aplicación se valoren conjuntamente.

Planeación

1) **Seleccione los aprendizajes esperados que evaluará.**
   A partir del análisis realizado anteriormente, ¿cuáles son los aprendizajes que pretendo valorar con el proyecto?

   Consejo 1: Esta estrategia puede emplearla durante un curso, o como producto final de éste.

2) **Defina el tipo de evaluación que realizará.**
   ¿Esta evaluación tiene un fin diagnóstico, formativo o sumativo?
3) **Establezca quiénes participarán en la evaluación.**
   ¿Se realizará una autoevaluación, heteroevaluación o coevaluación? Considere que puede emplear una o todas las anteriores.
4) **Especifique el tipo de proyecto.**
   De acuerdo con la naturaleza de los objetivos, el tiempo y los recursos con los que dispone, seleccione el tipo de proyecto que se adecue a sus necesidades.
5) **Puntualice el producto que utilizará.**
   Escoja un producto o productos que permitan valorar el objetivo que se planteó y que esté acorde con el tema, éstos pueden ser reporte, tesis, artículo, herramientas de sistematización de la información, diario de campo (bitácora), objeto, diseño, software, un portafolio de arte, examen oral (*viva*), póster, exposiciones o presentaciones.

   Consejo 2: Tome en cuenta que la elaboración y la evaluación implican actividades dentro y fuera de clase.

6) **Determine si será por equipo o individual.**
   ¿La valoración se realizará a nivel individual o grupal? En caso que el proyecto se desarrolle en equipo, ¿la calificación será individual o grupal?
7) **Precise con qué criterios evaluará el proyecto.**
   ¿Qué aspectos evaluará con este proyecto? ¿Cuáles serán los indicadores observables? ¿Qué peso le otorgará a cada indicador?

   Consejo 3: Los instrumentos se adecúan a la naturaleza de las actividades.

8) **Seleccione el instrumento que utilizará para la evaluación.**
¿Qué instrumento facilita la recopilación de información acerca del nivel de desempeño de los estudiantes y el logro de los aprendizajes?

Diseño

1) **Presente los criterios de valoración del proyecto.**
Explique y entregue por escrito los criterios de evaluación y el peso que asignará a cada uno, la manera en que se recabarán las evidencias, además de sus tiempos de entrega y características específicas.

Consejo 4: Los criterios los puede establecer el docente o realizar conjuntamente con los estudiantes.

2) **Delimite el tema del proyecto.**
Proponga un tema con contenido adecuado para el nivel educativo, este puede ser definido por el estudiante o en conjunto.
3) **Construya los instrumentos de evaluación.**
Para elaborarlos, considere que puede apoyarse de los manuales de cada instrumento, estos pueden utilizarse durante todo el desarrollo del proyecto o en una etapa.
4) **Asesore cada etapa del proyecto.**
De acuerdo con cada uno de los productos y su avance, valore el acompañamiento que requieran los estudiantes para lograr los objetivos.
5) **Promueva la revisión de los productos.**
Al revisar los productos los estudiantes podrán autoevaluarse al mismo tiempo que el docente lo evalúa. Puede recomendar el uso de esquemas del reporte o listas de cotejo.
6) **Informe los resultados de la evaluación.**
Además de proporcionar al estudiante una calificación, es importante que ésta venga acompañada de una retroalimentación para que mejore sus habilidades y conocimientos.
7) **Favorezca la comunicación pública de resultados.**
Promueva que los estudiantes comuniquen los resultados de sus proyectos con sus pares y colegas, por ejemplo, en ferias o festivales educativos.

Consejo 5: Considere la elaboración de posters, videos, exposiciones y otras formas innovadoras de presentación.

Revisión

- Se requiere analizar periódicamente los avances de cada una de las etapas del proyecto para retroalimentar a los estudiantes.
- Integre todas las evidencias del estudiante durante todo el proceso para una visión completa de su trabajo.

¿CÓMO LA APLICO?

- Es una estrategia que requiere de entregas parciales de productos, para valorar el avance y logro de los estudiantes y así poder sugerir acciones para mejorarla, por lo que es importante retroalimetar a los estudiantes de forma oportuna.
- Se puede evaluar la calidad de la presentación tanto oral como escrita, además de los productos que se entreguen en cada una de las etapas.
- Se requiere promover la colaboración entre los estudiantes, docentes y otras personas involucradas con el campo de conocimiento, para obtener consejos y otros puntos de vista con el fin de mejorar el producto.
- A continuación, se proponen algunos productos que se pueden solicitar a los estudiantes.
    - **Al inicio del proyecto**: plan de investigación, lista de recursos, análisis del problema, diseño del plan, prospecto y una secuencia de actividades.
    - **A la mitad del proyecto**: organización, bosquejo del proyecto, anteproyecto y revisión de criterios.
    - **A la terminación del proyecto**: síntesis, conclusiones, modelo final, documentación del producto, presentación final e informe del proyecto.
- Puede enfocarse en el contexto en el que vivimos, lo que lleva a una mayor motivación para el estudiante, ya que puede alinearse con sus preocupaciones, temas de interés y su vida cotidiana.
- Es importante considerar los recursos con los que cuentan los estudiantes para el desarrollo del producto, por ejemplo, materiales, espacios de trabajo, tiempo y acceso a la información, entre otros.

EL PROYECTO EN LA EVALUACIÓN A DISTANCIA

En esta modalidad es necesario que todas las actividades implicadas en la realización de un proyecto promuevan el trabajo colaborativo y la autorregulación en los estudiantes, ya que los beneficiará a lo largo de su carrera profesional. Para implementar esta estrategia, considere las siguientes recomendaciones que complementan la sección de "¿Cómo la diseño?" de este capítulo.

¿Cómo lo diseño a distancia?

- Analice el mapa curricular, perfil de egreso y/o programa de asignatura para seleccionar lo que evaluará en una modalidad a distancia. Considere los aprendizajes esperados fundamentales y transversales que le permitan diseñar un proyecto con actividades dinámicas y accesibles para sus estudiantes.
- Para definir la planeación del proyecto que se evaluará, se propone completar la siguiente tabla:

Tabla 1. Elementos para definir una evaluación a distancia de un proyecto

Elementos	Guía	Respuesta
Tema	Seleccione un tema fundamental de la asignatura que pueda evaluarse en línea mediante un proyecto. Considere si el tema tiene potencial para trabajar de forma multidisciplinaria con otras asignaturas.	
Aprendizajes esperados	Describa los conocimientos y/o habilidades que aprenderán en clase. Elija aquellos que desee valorar.	
Actividades	Plantee actividades que, además de evidenciar el aprendizaje esperado sean accesibles para usted y los estudiantes. También diseñe actividades multidisciplinarias con otras asignaturas. Favorezca actividades que impliquen trabajo en equipo. Incluya al menos una actividad, cuya finalidad sea sumativa, diagnóstica y formativa.	
Producto	Escoja un producto asequible a la modalidad a distancia. Considere su carga académica y la de sus estudiantes. Establezca la forma de entrega: Asíncrona (por ejemplo, envíos por correo electrónico o al Sistema de Gestión de Aprendizaje que esté utilizando). Sincrónica (por ejemplo, presentación en línea o envío de archivos en tiempo real en plataformas como Zoom).	
Duración	Para la planeación de las actividades y entregas de los productos, defina con cuántas semanas y sesiones efectivas cuenta. Esto le permitirá establecer: Fechas de entrega. Número de productos y/o actividades. Número de retroalimentaciones sincrónicas. Fechas de envío de retroalimentaciones asincrónicas.	
Recursos	Realice una búsqueda tecnológica de los recursos que podrían emplear los estudiantes. Considere en que punto del temario se encuentra para solicitar los productos. Incluya recursos que acerquen a los estudiantes al uso de ciertas herramientas que se requieren para la elaboración del proyecto. Por ejemplo, tutoriales para utilizar ciertos programas como: GIMP, Power Point, Excel, FreePik, AutoDraw, Draw.io, entre otros.	
Retroalimentación	Defina los aspectos a considerar durante la retroalimentación efectiva.	

- Si el proyecto es de carácter multidisciplinario, se sugiere que lo comente con las asignaturas relacionadas para tener un trabajo colaborativo y realizar una evaluación de forma colegiada.
- Establezca los criterios con que evaluará el proyecto y descríbalos de forma exhaustiva. Recuerde que estos se presentan en formato escrito en la modalidad a distancia.
- Incluya instrumentos de evaluación por pares y autoevaluación, orientados a identificar y potenciar habilidades necesarias para el aprendizaje autónomo y permanente.
- Al momento de redactar las instrucciones del proyecto, tome en cuenta la perspectiva del estudiante, incluya información que podría ser obvia para un experto en el tema, pero no para un aprendiz.
- Las instrucciones deberán presentarse en orden lógico, por actividades e indicando qué, cómo, cuándo y de qué manera se tiene que entregar.
- Planifique el tipo de asesoría que requiere cada actividad (tiempo, preparación, retroalimentación).
- En el caso de utilizar un Sistema de Gestión de Aprendizaje (LMS) defina las actividades y productos que serán calificados de forma automática, se sugiere consultar los manuales y tutoriales de LMS que utiliza su institución educativa para realizar esta actividad.

¿Cómo lo instrumento?

Integre las siguientes recomendaciones con el apartado "¿Cómo lo aplico?" contenido en este capítulo.

- Esta estrategia de modalidad a distancia requiere de asesorías, por lo tanto, se sugiere promover el trabajo de documentos en carpetas compartidas en *Google Drive* y solicite el acceso a estas carpetas. De esta manera usted podrá tener acceso a la forma en que trabajan los estudiantes y realizan sus aportaciones.[2]
- Brinde una serie de lineamientos que favorezcan el trabajo en equipo colaborativo y a distancia.
- En el caso que la evaluación se realice en equipo solicite un apartado en el reporte escrito que indique la contribución de cada uno de los estudiantes.
- Utilice las sesiones de retroalimentación sincrónicas para enfocar el proyecto y proponer preguntas que permitan avanzar a los estudiantes en el logro de los objetivos de la evaluación.
- Promueva la aplicación de autoevaluaciones para conocer la experiencia de los estudiantes en el uso de nuevas plataformas y saber en qué áreas se puede mejorar.
- Utilice los recursos que propondrá a los estudiantes, lo que le permitirá valorar su pertinencia y accesibilidad.

---

2 Canal Academy pop. (26 de abril del 2021) Cómo CREAR SITIO WEB para un ABP o APRENDIZAJE BASADO EN PROYECTO - Google Site desde cero 2021 [Archivo de Vídeo]. Youtube https://www.youtube.com/watch?v=hOTVisMdgaY&t=0s

- Establezca un medio de comunicación asíncrono en donde los estudiantes puedan expresarle sus dudas. Por ejemplo, puede proporcionar un correo electrónico, incluir un foro público en la LMS (*Moodle*) en que esté alojado el curso en línea, crear un grupo privado de *Facebook* o un blog (*tumblr* o *blogger*). Es importante que, al momento de presentar este medio, se indique en cuánto tiempo se dará respuesta y la fecha de cierre de este tipo de comunicación.
- Es importante tener una plática con los estudiantes acerca de la integridad académica y el plagio, ya que como tienen acceso a más fuentes electrónicas, se puede prestar a este tipo de conductas no deseadas.
- Para ver más recomendaciones, se sugiere ver el siguiente video: Canal GrupoEducare (14 de julio del 2020) ¿Cómo desarrollar proyectos en la educación a distancia? [Archivo de Vídeo]. Youtube. https://www.youtube.com/watch?v=PYZLbk8lQ3s&list=PL7Fputt7rzTtb3fqORQrpHuMetxI6Eq4l&index=3

¿CÓMO LO ANALIZO?

- Los aprendizajes esperados deben reflejarse en los criterios de evaluación, éstos permitirán al estudiante conocer qué es lo que se espera que desarrolle y que aprenda durante el curso.
- Es necesario percatarse de que cada trabajo o entrega refleje lo visto en el curso, que sea original y creativo, además de que se pueda observar el desarrollo del pensamiento con cada una de las entregas.
- Los productos de los proyectos pueden ser evaluados por el docente, por los pares o por los propios estudiantes, para analizar los resultados obtenidos en la elaboración de un proyecto, se puede utilizar una o varias de las siguientes técnicas:
    - **Registro de la presentación del resultado:** puede ser registrada en formato video, y de acuerdo a los criterios establecidos fomentar la autoevaluación y la evaluación por pares.
    - **Guión de preguntas:** las preguntas indagan sobre la evolución del proyecto.
    - **Rúbrica:** seleccionar los criterios y el peso que se le dará a cada uno de ellos.
    - **Evaluación por pares:** valorar, a partir de una encuesta, los criterios que se tomaron en cuenta para la rúbrica.
    - **Autoevaluación:** elaborar una serie de preguntas que propicien la reflexión sobre el tema y el desempeño del estudiante.
    - **Informes:** pedir testimonios o avances sobre el progreso del grupo.
    - **Diario de trabajo:** escribir la bitácora de trabajo por proyecto y producto.
    - **Guía de observaciones:** es un registro de las observaciones realizadas a lo largo del curso, el cual puede ofrecer un diagnóstico para evaluar al estudiante.
- Es necesario asignar un peso específico a cada una de las etapas del proyecto de acuerdo con la complejidad del tema y a las situaciones que puedan surgir durante el desarrollo del proyecto. Por ejemplo, considerar el porcentaje de avance de una etapa para asignar una calificación.

- Es importante tener en cuenta que los proyectos tienden a tomar su propio rumbo, por eso es importante evaluarlos de acuerdo con la efectividad del proyecto conforme se desarrolla, encauzarlo de ser necesario y valorarlo cuando esté terminado.
- El análisis de resultados debe ayudar a comparar y valorar las diferencias de dominio de contenidos entre grupos, estudiantes o de un mismo estudiante en diferentes momentos del aprendizaje, para ello puede explorar las características de los Sistemas de Gestión de Aprendizaje como: *Google Classroom*, *Moodle*, *Edmodo*, entre otras, a las que algunas instituciones educativas tienen acceso, ya que todas tienen herramientas para ayudar a calificar y de esa forma poder concentrarse en las retroalimentaciones efectivas.

EJEMPLOS

1. Lista de Cotejo para evaluar un anteproyecto de investigación

Instrucciones: Marque la casilla que corresponda si los criterios para evaluar el anteproyecto de tu compañero cumplen (SÍ), no cumplen (NO) o no aplican (NO APLICA) tomando en cuenta cada descripción.

Alumno: **Título:**			
CRITERIOS	SÍ	NO	NO APLICA
La revisión de las fuentes de información			
es relevante para el problema.			
es crítica.			
presenta un orden lógico.			
se vincula con el proyecto.			
El problema del proyecto			
se expone claramente.			
se plantea como una pregunta.			
se relaciona con la revisión de literatura.			
Los objetivos del proyecto			
se relacionan con el problema.			
son claros.			
son alcanzables.			

CRITERIOS	SÍ	NO	NO APLICA
La hipótesis			
está presente.			
está fundamentada.			
es clara.			
se relaciona con el problema.			
se relaciona con los objetivos.			
El método			
describe el diseño.			
contiene la información necesaria.			
es adecuado para responder el problema.			
responde la hipótesis.			
especifica la población y la muestra.			
describe las variables.			
describe los instrumentos.			
describe el análisis que se llevará a cabo.			
Los aspectos éticos			
se describen.			
Las referencias			
son citadas en todo el documento con formato solicitado. (Por ejemplo: APA, Chicago, Vancouver, entre otros)			
son actuales (últimos 5 años).			
El título			
permite identificar el problema de investigación y las variables.			
es conciso.			
OBSERVACIONES:			

Fuente: Elaboración propia

2. Escala valorativa de un proyecto

La presente escala valorativa está destinada a evaluar un trabajo de investigación en su primera etapa, con el fin de mejorar cada una de las secciones.

Título del proyecto:	
Alumno:	

Instrucciones: Señale la puntuación que corresponda según su criterio.
Utilice la siguiente escala:

5 puntos Excelente	4 puntos Muy Bien	3 puntos Bien	2 puntos Suficiente	1 punto Deficiente	0 puntos No lo presentó	N/A No aplica

CRITERIOS PARA EVALUACIÓN DEL PROYECTO							
1. Revisión crítica de la literatura y elaboración del marco teórico y conceptual							
1.1 Presenta una síntesis del contexto general (nacional y mundial) en el cual se ubica el tema del proyecto propuesto.	5	4	3	2	1	0	N/A
1.2 Presenta el estado actual del conocimiento del problema, las brechas que existen y el vacío que se pretende llenar con el proyecto.	5	4	3	2	1	0	N/A
1.3 Describe el/los modelo(s) conceptual(es) en que se realizará el trabajo.	5	4	3	2	1	0	N/A
1.4 Las referencias bibliográficas que documentan el problema son pertinentes y actualizadas.	5	4	3	2	1	0	N/A
2. Planteamiento del problema y justificación							
2.1 La propuesta presenta claramente un problema delimitado y concreto, o una situación a cuya solución, entendimiento o comprensión, contribuirá con la realización del proyecto de investigación.	5	4	3	2	1	0	N/A
2.2 Describe la relevancia, trascendencia y magnitud del problema y el proyecto.	5	4	3	2	1	0	N/A
2.3 La pregunta de investigación es clara y completa.	5	4	3	2	1	0	N/A
3. Objetivos de la investigación							
3.1 Los objetivos son definidos con claridad.	5	4	3	2	1	0	N/A
3.2 Los objetivos son congruentes con el planteamiento del problema y las preguntas y/o hipótesis que se pretenden resolver.	5	4	3	2	1	0	N/A
3.3 Los objetivos tienen sustento en el modelo conceptual.	5	4	3	2	1	0	N/A
3.4 Los objetivos son alcanzables con la metodología propuesta.	5	4	3	2	1	0	N/A

4. Hipótesis o Expectativas teórico-metodológicas							
4.1 Las hipótesis o expectativas del proyecto se plantean con claridad.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
4.2 Las hipótesis o expectativas son congruentes con el planteamiento del problema y los objetivos.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
4.3 El proyecto permite explorar a profundidad la problemática planteada (estudios cuantitativos, cualitativos o mixtos).	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
5. Diseño de investigación							
5.1 La metodología del diseño es consistente con el marco teórico y modelo conceptual.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
5.2 Describe el criterio para la selección de la población y/o muestra.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
5.3 Las técnicas de muestreo son apropiadas.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
5.4 Las variables se definen conceptualmente de manera apropiada.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
5.5 Las variables se definen operacionalmente de manera apropiada para su medición y/o interpretación.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
5.6 Identifica con claridad las estrategias metodológicas.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
5.7 Los instrumentos propuestos recaban la información necesaria para responder el problema.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
5.8 Las técnicas de análisis son apropiadas para la metodología seleccionada.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
6. Plan de análisis							
6.1 Presenta una propuesta inicial de plan de análisis de los datos.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
6.2 Explica cómo se alcanzarán los objetivos propuestos.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
6.3 Presenta los recursos propuestos de forma detallada y la factibilidad del trabajo.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
6.4 Presenta una relación razonable entre los rubros, objetivos, metodología y duración del proyecto.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
7. Aspectos éticos							
7.1 Considera las implicaciones éticas del proyecto.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
8. Cronograma de actividades							
8.1 Presenta adecuadamente la secuencia de actividades y tiempo previsto para su realización, para alcanzar los resultados esperados.	**5**	**4**	**3**	**2**	**1**	**0**	**N/A**
OBSERVACIONES:							

Fecha: ______________________

Adaptado de: Dr. Adrián Martínez González (2017)

3. Autoevaluación de un proyecto

Instrucciones: Con base en su proyecto, conteste el siguiente cuestionario, recuerde que las siguientes preguntas le ayudarán a conocer lo que ha aprendido durante el curso y a mejorar sus herramientas para desarrollar un proyecto.

Nombre:____Fecha:_________________

Título del proyecto:____

Núm. de Revisión:____

1. Como resultado de trabajar en este proyecto, he aprendido que mis fortalezas son

____

____

2. Como resultado de trabajar en este proyecto, he aprendido que me cuesta trabajo

____

____

3. Como resultado del trabajo en este proyecto, lo que aprendí del tema fue

____

____

4. Como resultado del trabajo, lo que aprendí sobre realizar un proyecto fue

____

____

5. Como resultado del trabajo, lo que aprendí sobre trabajar en equipo fue

____

____

6. Como resultado del trabajo, lo que aprendí sobre planificar actividades fue

____

____

7. Los comentarios que recibí tanto de mi profesor como de mis compañeros me han servido para

____

____

8. Como resultado de la retroalimentación, lo que realizaré es

____

____

Fuente: Elaboración propia

RECURSOS EN LÍNEA

Recurso	Descripción
Buck Institute for Education (BIE) https://www.bie.org	Es una página que ofrece apoyos para el uso de proyectos enfocados en la evaluación del aprendizaje.
Consort http://www.consort-statement.org/	Sitio que agrupa listas de cotejo para valorar distintos tipos de proyectos.
Prisma https://www.equator-network.org/reporting-guidelines/prisma/	Sitio que agrupa lista de cotejo y lineamientos para escribir y valorar reportes sistemáticos y metaanálisis.
Perfeccionamiento del Profesorado DGOIC – CEUCD Introducción al Aprendizaje Basado en Proyectos (ABP) https://www.youtube.com/watch?v=mtBHSNzFGOM	Video introductorio al ABP.
Aprendizaje basado en proyectos https://www3.gobiernodecanarias.org/medusa/ecoescuela/pedagotic/aprendizaje-basado-proyectos/	Descripción sucinta de la metodología de Aprendizaje Basado en Proyectos. Incluye una infografía con los principales datos.
Recursos TIC para la evaluación ONLINE https://www3.gobiernodecanarias.org/medusa/ecoescuela/pedagotic/herramientas-online-para-evaluar/	Esta guía presenta ideas clave sobre la evaluación formativa y un conjunto de fichas e infografía sobre recursos TIC que pueden ser utilizados para implementar diferentes instrumentos de evaluación.
Plagio académico y ética en la Tesis – UNAM https://www.youtube.com/watch?v=NmQ6UnlyRCk&t=326s	Vídeo que muestra distintos tipos de plagio académico, sobre la ética académica y propone una actividad.

CONCLUSIONES Y REFLEXIONES FINALES

- La evaluación basada en proyectos es interdisciplinaria, lo que permite que los estudiantes relacionen conocimientos y contenidos de diferentes disciplinas para la solución del problema planteado o la elaboración de un producto.
- Permite reconocer si un estudiante tiene las actitudes y el interés por ser investigador.
- Es importante considerar que el tema del proyecto sea adecuado e interesante para los estudiantes para que se comprometan y se esfuercen por concluirlo.
- Recopilar proyectos anteriores que puedan servir de ejemplo a los estudiantes para darles una visión realista, que sirva como fuente de inspiración.
- Utilice instrumentos como el portafolio, ya que recopila la información del proceso que ha llevado a cabo el estudiante durante el desarrollo del proyecto.

REFERENCIAS

Ahumada-Acevedo, P. (2005) *La Evaluación auténtica: un sistema para la obtención de evidencias y vivencias de los aprendizajes. Perspectiva Educacional, formación de profesores*, 11-24. https://www.redalyc.org/articulo.oa?id=333329100002

Brown, G., Bull, J. y Pedlebury, M. (1997). Assessing projects. En *Assessing student learning in higher education*. Nueva York, Estados Unidos: Routledge.

Buck Institute for Education (6 de septiembre del 2021) PBL Project Based Learning. https://www.pblworks.org/

Herrán, A. de la (2011). Técnicas didácticas para una enseñanza más formativa. En N. Álvarez y R. Cardoso (Coords.). *Estrategias y metodologías para la formación del estudiante en la actualidad*. Camagüey, Cuba: Universidad de Camagüey.

Instituto Tecnológico y de Estudios Superiores de Monterrey, T. (2000). *El método de proyectos como técnica didáctica*. Dirección de Investigación y Desarrollo Educativo. Vicerrectoría Académica. Instituto Tecnológico y de Estudios Superiores de Monterrey. Monterrey, México: DR ITESM.

Ko, S., y Rossen, S. (2017). Course design and development. En: Teaching online: *A practical guide* (pp. 33-91). Routledge.

López, B. y Hinojosa, E. (2000). *Evaluación del aprendizaje. Alternativas y nuevos desarrollos*. Ciudad de México, México: Editorial Trillas. http://sgpwe.izt.uam.mx/files/users/virtuami/file/ext/gestion_evaluacion_actv_ext_tecnicasalternativas.pdf

Palacios B., Sánchez, M. C. y Gutiérrez A. (2013). *Evaluar la calidad en la investigación cualitativa: Guías o checklists*. http://uvadoc.uva.es/handle/10324/3014

The State of Queensland. The Office of the Queensland Studies Authority (2002). *Sourcebook module. Introducing robotics.* https://www.qcaa.qld.edu.au/downloads/p_10/kla_tech_sbm_603.pdf

Capítulo 31

INVESTIGACIÓN

María Juliana Londoño Cárdenas, María Elena Pérez Rivera, Adrián Martínez González

"En algún lugar, algo increíble está esperando ser conocido."
Carl Sagan

INTRODUCCIÓN

Una de las funciones principales que debe realizar el profesor dentro de los escenarios educativos es evaluar los aprendizajes de los estudiantes.

Es importante reconocer que la evaluación no se debe centrar solamente en los conocimientos, sino también en su aplicación.

La evaluación basada en la investigación permite valorar no solo el contenido sino distintas formas de pensamiento, logrando potenciar el proceso enseñanza-aprendizaje tanto en la modalidad presencial como a distancia.

Esta estrategia de evaluación es una de las más empleadas por los docentes para evaluar el desempeño de los estudiantes, ya que promueve el pensamiento crítico, la solución de problemas, una estructuración lógica de actividades, así como actitudes positivas hacia el trabajo escolar, innovación y creatividad. Le permite al docente y al estudiante el análisis del proceso de aprendizaje que realiza al elaborar la investigación. El docente observa quién es, cómo es y qué puede llegar a hacer el estudiante, pensando en cómo apoyarlo en su formación y a su vez el estudiante puede autoevaluarse, reconociendo qué aprendizajes ha adquirido al realizarla.

Al igual que en otras estrategias de evaluación del y para el aprendizaje, el papel del docente es muy importante ya que no solo debe revisar el resultado sino acompañar al estudiante durante el proceso, actuando como facilitador y guía. El docente y el estudiante aprenden e investigan al mismo tiempo. El docente debe evaluar basándose en el desempeño, el proceso y los productos que entreguen los estudiantes, y a su vez ofrecerles una retroalimentación eficaz y oportuna.

Esta estrategia es de las más utilizadas durante el periodo de pandemia ya que las actividades relacionadas con una investigación, permiten que se realicen de manera asíncrona y

al ser en línea, los recursos tecnológicos utilizados son accesibles y familiares para los estudiantes, lo que facilita su implementación (Peimani y Kamalipour, 2021).

Este capítulo pretende presentar, de manera sencilla, información que le permita implementar esta estrategia que forma parte de los métodos cualitativos de evaluación. Se organiza en ocho secciones. En la primera se define qué es una evaluación basada en una investigación, en la segunda se detallan los pasos para diseñarla, en la tercera se recomienda cómo aplicarla, en la cuarta se propone su utilización en la modalidad a distancia, la quinta sugiere cómo analizar sus resultados, en la sexta se incluyen tres ejemplos, en la séptima se presentan recursos en línea y en la octava las conclusiones.

¿QUÉ ES?

- Es un proceso sistemático, cíclico, complejo, que permite valorar el desarrollo del razonamiento del estudiante para la solución de problemas en el proceso de enseñanza-aprendizaje.
- Es una estrategia que sirve tanto para el aprendizaje como para la evaluación, ya que se puede realizar en etapas que se valoran y en las que se puede hacer una retroalimentación a los estudiantes para que reformulen y avancen en el proceso de investigación (Verdejo, Encina y Trigos, 2011). Es utilizada en todos los niveles educativos, especialmente en el caso de los posgrados es la estrategia más utilizada.
- Puede ser una actividad dirigida a la solución de problemas en la que la teoría y la práctica deben estar relacionadas teniendo unos objetivos y una metodología a seguir. El proceso y el producto de la investigación sirve para reconocer la manera en que los estudiantes aplican lo aprendido para resolver un cuestionamiento de forma sistemática.
- Existen **tres etapas** en la investigación:
  - **Planificación**: se definen el problema, propósito, antecedentes y diseño, en un artículo serían las secciones de introducción y método.
  - **Ejecución:** se presentan los resultados y el análisis, las secciones que pertenecen a esta etapa son: resultados, discusión y conclusiones.
  - **Comunicación de los resultados:** se pueden difundir los principales hallazgos de acuerdo al nivel educativo y al tipo de profundidad en que se desarrolle el tema, los cuales pueden ser un reporte, un póster, una exposición, un portafolio y un artículo, entre otros.

- Es posible investigar desde diferentes perspectivas; considerando ya sea objetivos o tomando en cuenta tipos de datos, procedimientos o métodos para obtenerlos. Como se explican a continuación:
- El objetivo y propósito con el que se realizan pueden ser:
  - **Teórico:** busca generar conocimientos que permitan entender, explicar y comprender ciertos fenómenos.
  - **Aplicado:** busca la aplicación o la utilización de los conocimientos que se adquieren.

- El nivel de profundización en el objeto de estudio puede ser:
  - **Exploratoria:** examina un tema o problema de investigación poco estudiado del cual se tienen muchas dudas o no se ha abordado antes.
  - **Descriptiva:** pretenden medir o recoger información de manera independiente o conjunta sobre los conceptos o las variables a las que se refiere.
  - **Causal:** su finalidad es conocer la relación o grado de asociación que existe entre dos o más conceptos, categorías o variables en un contexto en particular.
  - **Explicativa:** como su nombre lo indica explica por qué ocurre un fenómeno y en qué condiciones se manifiesta o por qué se relacionan, dos o más variables.

- El tipo de datos empleados puede ser:
  - **Cualitativa:** emplea la recolección de datos sin medición numérica para descubrir o afinar preguntas de investigación en el proceso de interpretación.
  - **Cuantitativa:** usa la recolección de datos para probar hipótesis con base en la medición numérica, y utiliza el análisis estadístico para establecer patrones de comportamiento y probar teorías.
  - **Mixta:** integra una interpretación de datos cuantitativos y cualitativos para realizar inferencias de toda la información y lograr un mayor entendimiento del fenómeno bajo estudio.

- El periodo temporal en que se realiza la medición de las variables puede ser:
  - **Longitudinal:** analizar variables a través del tiempo en puntos o periodos para hacer inferencias respecto al cambio, sus determinantes y consecuencias.
  - **Transversal:** recolectar datos de las variables en un solo momento en un tiempo único.

- El desarrollo de una investigación pone en juego los conocimientos de diferentes asignaturas o módulos, por lo que permite promover el trabajo inter y multidisciplinario entre profesores y estudiantes.
- El proceso y el producto de la investigación sirven para reconocer la manera en que los estudiantes aplican lo aprendido para resolver un cuestionamiento de forma sistemática.
- La evaluación por pares y la autoevaluación promueven la reflexión y el pensamiento crítico en el estudiante.

¿CÓMO LA DISEÑO?

Para una evaluación basada en investigación, es necesario realizar un análisis del mapa curricular, del perfil de egreso y del curso, módulo o programa de asignatura para considerar qué se quiere investigar, el por qué o para qué para darle sentido, tanto a la actividad de investigación como a su evaluación y, al valorar el cómo investigar, podemos generar acciones sistemáticas que puedan contribuir a la formación del estudiante.

Planeación

1) **Escoja los aprendizajes esperados que se evaluarán.**
   De acuerdo con el análisis realizado anteriormente, ¿qué aprendizajes se adecuan para evaluar mediante una investigación?

   Consejo 1. Tome en cuenta que la elaboración y la evaluación implican actividades dentro y fuera de clase.

2) **Establezca el tipo de evaluación que realizará.**
   ¿Esta evaluación tiene un fin diagnóstico, formativo o sumativo? Considere que dentro de la evaluación puede utilizar actividades con otros fines al que eligió.
3) **Determine quiénes participarán en la evaluación.**
   ¿Se realizará una autoevaluación, heteroevaluación o coevaluación? Considere que puede emplear una o todas las anteriores.
4) **Defina el proceso de investigación a evaluar.**
   Se puede evaluar un elemento de la etapa, como el planteamiento de un problema, también una etapa, como la presentación del análisis de resultados (ejecución); o bien, como un conjunto al evaluar un artículo o una tesis.
5) **Especifique con qué criterios evaluará la investigación.**
   ¿Qué aprendizajes esperados evaluará con esta investigación? ¿Cuáles serán los indicadores observables? ¿Qué peso le otorgará a cada indicador?
6) **Seleccione el instrumento que utilizará para la evaluación.**
   Para evaluar puede apoyarse de distintos instrumentos de evaluación como rúbrica, lista de cotejo, entre otros.

Diseño

1) **Explique claramente los criterios a evaluar.**
   Los criterios deberán estar acorde a su nivel educativo y entregarse por escrito. El documento deberá contener los indicadores de evaluación, el peso asignado, los tiempos de entrega por actividad o etapa y una descripción precisa de la elaboración y entrega de cada producto.
2) **Delimite el problema de investigación.**
   Seleccione un problema de investigación con contenido relevante para el estudiante y para su curso, módulo o asignatura, considere que esta selección la puede realizar junto con él.

   Consejo 2. Puede plantear el problema mediante un caso, una pregunta de investigación, un cuestionamiento directo o la resolución de una situación para que el estudiante lo identifique.

3) **Asesore el planteamiento del problema de investigación.**
Garantice que el planteamiento del problema de investigación sea relevante, pertinente, significativo y alcanzable para sus estudiantes, según su nivel de habilidades y conocimientos.
4) **Elabore los instrumentos de evaluación.**
De acuerdo con las necesidades de evaluación construya o seleccione un instrumento que le permita evidenciar los aprendizajes esperados; tenga en cuenta que puede apoyarse de los manuales de cada instrumento. Éstos pueden implementarse al principio, durante o al final del desarrollo de la investigación.
5) **Programe sesiones de retroalimentación.**
Deberán desarrollarse en un ambiente respetuoso que propicie que los estudiantes discutan entre ellos sus avances o dificultades.

Consejo 3. Precise la manera en que registrará sus retroalimentaciones.

6) **Planifique un tiempo de presentación de los trabajos.**
Considere una breve presentación del trabajo final de cada uno de los estudiantes, ya sea una exposición, un póster, un video, un debate, etcétera, para promover la discusión y el reconocimiento entre los estudiantes.
7) **Comunique los resultados de la evaluación.**
Asegure que los resultados se encuentren acompañados de comentarios que contribuyan a la comprensión de los temas y habilidades implicados en el desarrollo de uno o varios componentes de la investigación.

Revisión

- Es necesario analizar periódicamente los avances de cada una de las etapas de la investigación para retroalimentar a los estudiantes.
- Se debe tener en cuenta que es necesario integrar todas las evidencias de trabajo del estudiante, en caso contrario es posible que tenga una visión fragmentada de lo que pretende evaluar.
- Se sugiere compartir la implementación de esta estrategia con otros profesores para discutir y reflexionar sobre su experiencia y así mejorarla.
- Si dentro de sus actividades considera una revisión bibliográfica, solicite una entrega del material consultado comentado (referencia, palabras clave, síntesis del contenido y utilidad).

¿CÓMO LA APLICO?

- La evaluación basada en investigación requiere que se incorporen diferentes evidencias del desempeño de los estudiantes durante una o varias etapas de la investigación.
- La retroalimentación oportuna a los estudiantes conduce a llevar a cabo ajustes en cualquiera de las etapas que se solicite para realizar la investigación.

- La recopilación de evidencias de todo el proceso, no solo ofrece a los estudiantes retroalimentación para reflexionar, también constituye la información para valorar y tomar decisiones sobre las calificaciones que se otorgarán a los estudiantes.
- La retroalimentación y las sugerencias dadas a los estudiantes se pueden compartir con todo el grupo, lo que aumenta las posibilidades de mejorar el desempeño del grupo.
- Es importante que las técnicas de análisis y síntesis que utilicen los estudiantes como mapas mentales, cuadros sinópticos, síntesis, tablas, gráficos, entre otras, estén incorporadas en los trabajos que presenten.
- Se debe considerar bibliografía básica tanto de la metodología para llevar a cabo una investigación como del problema planteado para sugerir a los estudiantes.
- Se sugiere tomar en cuenta la disponibilidad de recursos de los estudiantes para el desarrollo de una o varias etapas, por ejemplo:
  - **El tiempo:** Puede ser desde una hora hasta abarcar todo el año escolar.
  - **El acceso a la información:** Uso de recursos digitales, acceso a bibliotecas, entrevistas, visitas a museos, entre otros.
  - **Los recursos económicos:** Tomar en cuenta si es necesario el uso de materiales para su elaboración, o si es necesario realizar entrevistas o búsquedas de información que requieran traslados.
  - **Los espacios de trabajo:** Considerar si es necesario el uso de laboratorios, si puede realizarse en casa, en bibliotecas, en el salón de clases, entre otros.

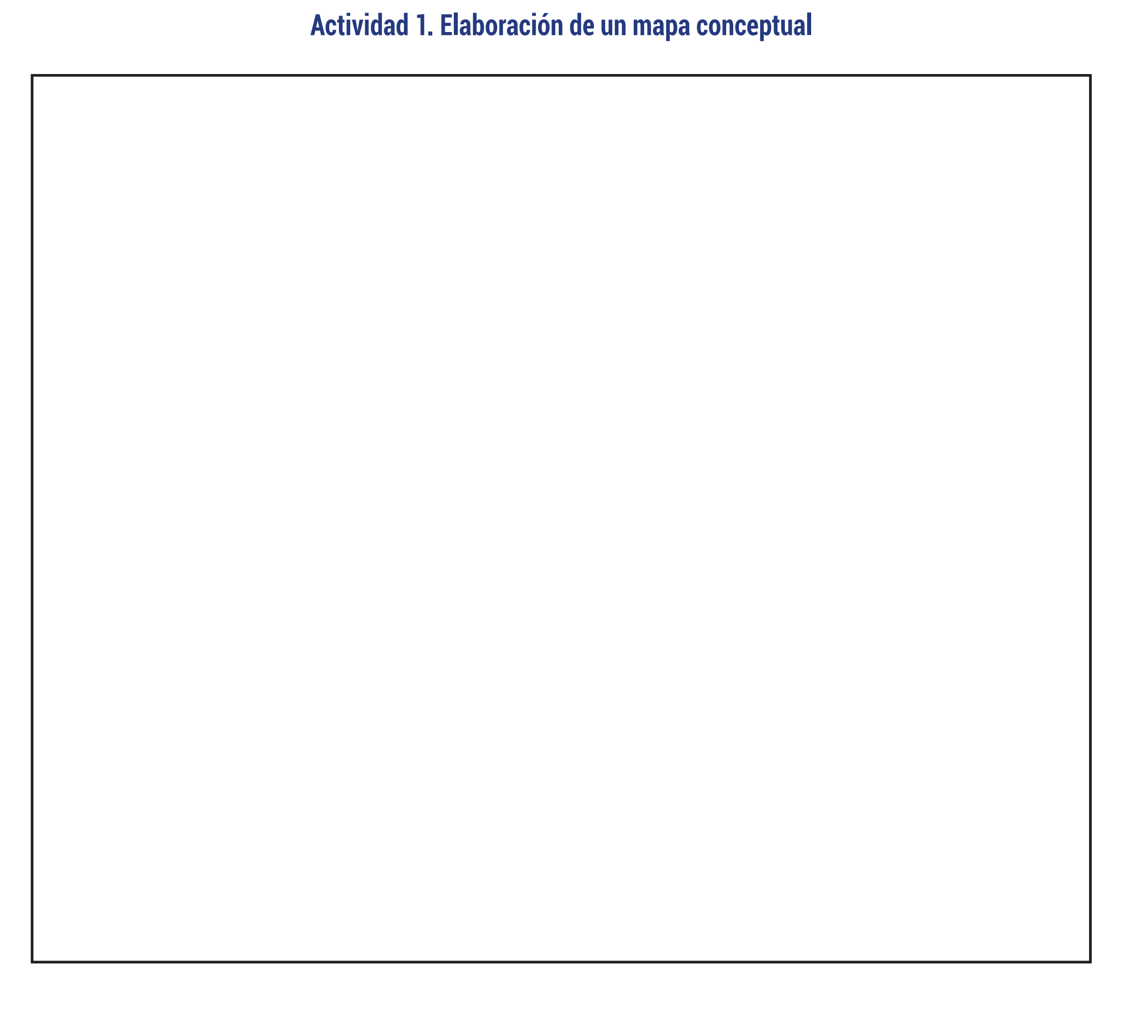
LA INVESTIGACIÓN EN LA EVALUACIÓN A DISTANCIA

En el diseño de una evaluación basada en investigación en la modalidad a distancia, tome en cuenta los elementos mencionados en la sección de "¿Cómo la diseño?" y compleméntelos con el siguiente apartado.

¿Cómo la diseño a distancia?

- Elabore un cronograma de avances que considere la sección de la investigación que tienen que presentar los estudiantes. Recuerde que las sesiones sincrónicas y asincrónicas deben alinearse con los temas propuestos en la investigación.
- Dependiendo de las situaciones personales en las que se encuentren sus estudiantes, sea flexible en cuanto fechas de entrega y actividades.
- Programe una sesión con la biblioteca de su institución educativa para que los estudiantes conozcan todos los recursos en línea que tienen disponibles de manera gratuita. En el caso que no sea posible esta sesión, comparta la página de los servicios de la biblioteca digital que presente manuales o instructivos.
- Tome en cuenta los cursos previos de sus estudiantes en referencia a la metodología en investigación para poder delimitar el alcance y orientación. Para ello, le recomendamos la siguiente actividad:

Actividad 1. Elaboración de un mapa conceptual



Instrucciones: En el recuadro elabore un mapa conceptual que relacione las asignaturas previas y actuales que tiene su estudiante lo que le servirá para identificar en donde tiene que incluir actividades para reforzar, desarrollar y concluir con la investigación; o con que asignaturas se podría trabajar en conjunto.

- Redacte las instrucciones siguiendo un orden lógico y coherente con la metodología de investigación que se haya seleccionado.
- Incorpore en las instrucciones un apartado de referencias bibliográficas agrupadas por actividad, producto, software o programas, que ayudarán al estudiante menos experimentado en algún tema específico.
- Incluya un apartado de discusión en el cual el estudiante relacione el contenido específico visto en las sesiones sincrónicas y asincrónicas con el tema de investigación.
- Limite la longitud de los documentos, reportes o textos a los estudiantes, ya que lo beneficiará al momento de revisar y brindar retroalimentación; y a sus estudiantes les permitirá desarrollar habilidades de síntesis.

- Diseñe o seleccione una rúbrica que valore aspectos de redacción y gramática para que acompañe sus retroalimentaciones asíncronas, de esta manera puede orientar los comentarios y observaciones hacia el contenido del proyecto. (Ko y Rossen, 2017).
- Programe una sesión sincrónica de presentación de la investigación en el que se plantee preguntas de discusión que enriquezcan la comprensión del tema y permita valorar la originalidad del trabajo.

¿Cómo lo instrumento?

Para incluir este tipo de estrategia en una evaluación a distancia, considere los puntos expuestos en la sección "¿Cómo la aplico?" de este capítulo en conjunto con las siguientes sugerencias.

- Utilice sitios web, plataformas, programas o softwares que sean pertinentes y accesibles a sus estudiantes para el desarrollo de los productos o actividades de la evaluación. En caso de que usted o sus estudiantes no tengan el dominio suficiente en el uso de los recursos tecnológicos pueden auxiliarse de tutoriales o plataformas guía para ello.
- Aloje en un programa de almacenamiento de archivos o datos (*drive, dropbox,* blog) comentarios, ideas o preguntas frecuentes, que surjan de las retroalimentaciones sincrónicas y que puedan utilizar los estudiantes para apoyar su aprendizaje.
- Desarrolle una carpeta en línea de material de consulta y recursos gratuitos que pueda apoyar a los estudiantes como libros, videos en *YouTube*, charlas *TED*, entre otros.
- Presentar junto con las instrucciones de la investigación un video que hable acerca de los peligros de plagiar información, se sugiere mostrar el siguiente video: Canal Prepa en línea (15 de diciembre del 2017). *Peligros del plagio* [Archivo de Vídeo] YouTube. https://www.youtube.com/watch?v=k9kJRLeQz50
- Si dentro de sus actividades considera una revisión bibliográfica, podría solicitar una entrega del material consultado comentado.
- En caso de solicitar mapas mentales, cuadros sinópticos, tablas, gráficos, puede sugerir el uso de plataformas como: *Draw.io*, *Coggle*, *Miro*, *Lucidchart*[1] Antes de presentar las plataformas, experimente con ellas y de ser necesario busque apoyo o tutoriales en línea.

¿CÓMO LO ANALIZO?

- Los resultados de una evaluación basada en investigación se apoyan en el uso de varios instrumentos, los cuales permiten obtener fuentes de información sobre el desempeño de los estudiantes.
- Los criterios de evaluación, que puede utilizar para analizar la información de una etapa o las tres, se pueden encontrar en la rúbrica de la sección de ejemplos. No es obligatorio seguir los criterios mostrados, estos pueden ser tomados como una guía para

[1] Lucidchart Español (14 de junio del 2018) Cómo Crear un Mapa Conceptual. [Archivo de Vídeo] YouTube. https://www.youtube.com/watch?v=q4QywyLYWrA&t=6s

realizar cualquier herramienta de evaluación que ayude a evaluar la etapa o etapas de investigación.

- De los instrumentos que se presentan a continuación, los primeros pueden servir tanto para evaluar lo que los estudiantes aprendieron en los escenarios educativos como para otorgar una calificación, los dos últimos solamente se recomiendan como una herramienta para evaluar de manera formativa, rescatar aquellos elementos que puedan ayudar a fortalecer el proceso de aprendizaje del estudiante.
  - **Rúbrica**: Criterios y peso de cada rubro para evaluar el trabajo.
  - **Lista de cotejo:** Presencia o ausencia de criterios en el trabajo.
  - **Diario de trabajo**: Bitácora de la investigación.
  - **Portafolio físico o portafolio en línea**: Recolección de evidencias, por ejemplo: ensayos, diarios, resolución de problemas, protocolos, entre otros.
  - **Evaluación por pares**: Resolver un cuestionario acerca del trabajo del otro proporcionando información que pueda ser relevante para fortalecer el aprendizaje.
  - **Autoevaluación**: Elaborar una serie de preguntas para el estudiante, que propicie su reflexión acerca del tema.

- Cada una de las herramientas proporcionarán observaciones significativas que servirán para poder orientar la retroalimentación que le dará al estudiante, la cual puede ser, sugerir artículos, aclarar conceptos, entre otros. Estos elementos pueden ayudarle en el desarrollo del trabajo.
- En caso de ser evaluación cuantitativa, puede tomar en cuenta el porcentaje de avance de la investigación o de la etapa que haya definido al inicio de la planeación de la evaluación para otorgar una calificación. Por ejemplo, al elaborar una lista de cotejo o una rúbrica podrá definir el porcentaje de avance suficiente para aprobar el curso ya sea seleccionando una sección o varias.

EJEMPLOS

1. Instrumento de autoevaluación de la sección de método

Objetivo. Valorar la sección de método de tu investigación.
Instrucciones. Considera la información que se ha trabajado a lo largo del curso y contesta lo siguiente.
Nombre del documento.

1. La sección de método presenta:	Sí	No
La manera en que se organiza.	☐	☐
El diseño de investigación.	☐	☐

La población.	☐	☐
La muestra.	☐	☐
Las variables del estudio.	☐	☐
Los instrumentos que se van a utilizar.	☐	☐
El proceso que se llevará a cabo para la recopilación de los datos.	☐	☐
La sistematización de los datos.	☐	☐
El análisis de datos.	☐	☐
Los aspectos éticos de la investigación.	☐	☐
2. Teniendo en cuenta el formato y estilo utilizado, el documento	**Sí**	**No**
Es claro.	☐	☐
Es preciso.	☐	☐
Tiene un estilo de escritura apropiado.	☐	☐
Tiene errores ortográficos.	☐	☐
Tiene errores sintácticos.	☐	☐
3. ¿Consideras que alguna de las ideas presentadas es relevante para tu formación académica?		
¿Por qué?		
4. ¿Qué apartado del método consideras que se desarrolló de mejor forma?		
¿Por qué?		
5. ¿Qué apartado del método consideras que es susceptible de mejorar?		
¿Por qué?		
6. ¿Qué consejo le darías al profesor para mejorar el desarrollo de la metodología?		

Fuente: Elaboración propia

2. Rúbrica para valorar un informe de investigación

Objetivo: Evalúa si el documento reporta el proceso realizado durante la investigación. Para esto, se valora qué tan completa se encuentra cada una de las secciones, esto se refiere a si los elementos esperados están presentes de manera cabal.

Instrucciones: Marque el nivel en que se logró completar cada sección considerando el criterio y la descripción del nivel de desempeño propuesto. En caso de que una sección o un criterio no aplique al tipo de reporte por favor señale con una marca la columna "No aplica".

Sección	Criterio	Muy completo	Completo	Incompleto	Muy incompleto	No aplica
Introducción	Planteamiento del problema	Define claramente el problema y el propósito de la investigación. Sustenta la justificación.	Define el problema y el propósito de la investigación. No está claramente sustentado con evidencia relevante.	Define el problema y el propósito de manera ambigua. Su justificación presenta confusión.	El problema no es claro y no se presenta propósito ni justificación.	
	Pregunta de investigación	La pregunta está planteada de manera clara y se relaciona con el problema.	La pregunta se relaciona con el problema.	La pregunta necesita mejorarse para relacionarse con el problema.	La pregunta causa confusión y no se relaciona con el problema.	
	Antecedentes	Las fuentes de información tienen mucha relación con el problema bajo estudio.	Las fuentes de información tienen relación con el problema bajo estudio.	Las fuentes de información tienen una relación limitada con el problema bajo estudio.	Las fuentes de información no tienen relación con el problema bajo estudio.	
	Objetivos de investigación	Los objetivos están planteados de manera clara y se relacionan con el problema.	Los objetivos se plantean y se relacionan con el problema.	Los objetivos necesitan reformularse para relacionarse con el problema.	Los objetivos no se relacionan con el problema.	
	Hipótesis	La hipótesis está formulada como posible respuesta al problema.	La hipótesis se relaciona con el problema, pero no está formulada como respuesta al problema.	La hipótesis no se relaciona con el problema.	La hipótesis es confusa y no se relaciona con el problema.	

Método	Diseño de investigación	El diseño se presenta de manera detallada.	El diseño se presenta de forma parcial.	El diseño se presenta por partes.	El diseño no se presenta.	
	Población y muestra	Describe la población y el método de muestreo con claridad.	Describe parcialmente la población y el método de muestreo.	Describe la población, pero no el método de muestreo.	No describe ni la población ni el método de muestreo.	
	Variables del estudio	Se identifican todas las variables del estudio. Las variables fueron definidas de forma conceptual y operacional.	Algunas variables fueron definidas conceptual y operacionalmente.	Las variables no fueron definidas operacionalmente.	No se identifican las variables y no existe su definición conceptual y operacional.	
	Recopilación de datos	La recopilación de datos se explica de manera detallada.	La recopilación de datos se explica de manera parcial.	La recopilación de datos se explica de manera superficial.	La recopilación de datos no está explicada.	
	Descripción del análisis	El análisis es adecuado para contestar las preguntas o someter a prueba las hipótesis.	El análisis es específico para contestar las preguntas o someter a prueba las hipótesis.	El análisis presenta limitaciones para contestar las preguntas o someter a prueba las hipótesis.	El análisis no es adecuado para contestar las preguntas o someter a prueba las hipótesis.	
	Aspectos éticos	Se utilizan los protocolos para la protección de los sujetos y se describen con detalle.	Se utilizan los protocolos para la protección de los sujetos y se describen de manera limitada.	Se utilizan los protocolos para la protección de los sujetos, pero no son descritos.	No se utilizan los protocolos para la protección de los sujetos ni se describen.	
Resultados	Descripción	Señala exhaustivamente los resultados expresados en valores y/o en forma descriptiva.	Señala los resultados expresados en valores y/o en forma descriptiva.	Señala parcialmente los resultados expresados en valores y/o en forma descriptiva.	Ausencia de los resultados expresados en valores y/o en forma descriptiva.	
	Presentación	El uso de figuras o tablas permite presentar los resultados de manera clara.	El uso de figuras o tablas es adecuado al presentar los resultados.	Se presentan los resultados en figuras o tablas de forma confusa.	No se presentan los resultados de manera clara.	
	Congruencia	Los resultados corresponden totalmente con la pregunta de investigación y objetivos.	Los resultados corresponden con la pregunta de investigación y objetivos.	Los resultados corresponden parcialmente con la pregunta de investigación y objetivos.	Los resultados no corresponden con la pregunta de investigación y objetivos.	

Discusión y Conclusiones	Comparación con los estudios previos	Compara los resultados obtenidos en este estudio con estudios previos.	Compara de manera limitada los resultados obtenidos en este estudio con estudios previos.	Compara uno de los resultados obtenidos en este estudio con estudios previos.	No compara los resultados obtenidos en este estudio con estudios previos.	
	Pregunta de investigación y análisis	La pregunta de investigación y el análisis de los resultados se consideran claramente en las conclusiones.	La pregunta de investigación y el análisis de los resultados se consideran parcialmente en las conclusiones.	La pregunta de investigación y el análisis de los resultados se presentan con confusión en las conclusiones.	La pregunta de investigación y el análisis de resultados no se presentan en las conclusiones.	
	Recomendaciones	Las recomendaciones surgen de un análisis de las conclusiones.	Las recomendaciones surgen parcialmente de un análisis de las conclusiones.	Las recomendaciones no surgen de un análisis de las conclusiones.	No presenta recomendaciones.	
Bibliografía	Calidad de las fuentes de información	La mayoría de las fuentes de información son variadas y se citan correctamente.	Las fuentes de información pertenecen a un solo autor y se citan correctamente.	Las fuentes de información pertenecen a un solo autor y no se citan correctamente.	Las fuentes de información no son variadas y no se citan correctamente.	
Título	Título	Permite la identificación del problema de investigación y contiene las variables del estudio.	Permite la identificación del problema de investigación, pero no las variables de estudio.	Contiene las variables del estudio, pero no el problema de investigación.	No permite la identificación del problema de investigación ni las variables del estudio.	
Resumen	Resumen	Describe de forma estructurada, sucinta y clara los objetivos, el método, resultados y conclusiones.	Describe de forma clara pero no sucinta los objetivos, el método, resultados y conclusiones.	Describe de forma escueta pero no clara los objetivos, el método, resultados y conclusiones.	No describe los objetivos, el método, resultados y conclusiones.	
Estilo	Formato	Sigue las recomendaciones de estilo que se le indicaron.	Sigue parcialmente las recomendaciones de estilo que se le indicaron.	Sigue un estilo, pero no el que se le indicó.	No sigue las recomendaciones de estilo que se le indicaron.	

Fuente: Elaboración propia

3. Coevaluación de una investigación mediante una escala valorativa

Objetivo: Evaluar en qué medida se mostraron los resultados de la investigación en la presentación final.
Nombre del trabajo: ____

Instrucciones: De acuerdo con la siguiente escala, indica en qué medida se presentaron los resultados de la investigación de tu compañero.

	Escala valorativa				
	Totalmente en desacuerdo	En desacuerdo	Indeciso	De acuerdo	Totalmente de acuerdo
El estudiante					
Presenta la descripción del problema de forma precisa.					
Describe los resultados de manera concisa.					
Utiliza gráficas o tablas que son claras.					
Argumenta basándose en los resultados presentados.					
Concluye tomando en cuenta los resultados.					
Relaciona sus conclusiones con el problema de investigación.					
El tema de investigación					
Es relevante.					
Es interesante.					
Es innovador.					

Recomendaciones:

____

____

____

____

Fuente: Elaboración propia

RECURSOS EN LÍNEA

Recurso	Descripción
Consort http://www.consort-statement.org/	Sitio que agrupa listas de cotejo que servirán para valorar distintos tipos de investigación.
Prisma https://www.equator-network.org/reporting-guidelines/prisma/	Sitio que agrupa lista de cotejo y lineamientos para escribir y valorar reportes sistemáticos y metaanálisis.
The EQUATOR Network http://www.equator-network.org/library/spanish-resources-recursos-en-espanol/	Sitio que agrupa diferentes instrumentos para valorar diferentes tipos de investigaciones.
Buck Institute for Education https://www.pblworks.org/	Sitio sobre el aprendizaje basado en proyectos e investigaciones. Los recursos y publicaciones a pesar de estar orientados a educación básica brindan información relevante y pertinente para cualquier nivel académico.
El Plagio (y Cómo Evitarlo - Clase Taller UNAM) https://www.youtube.com/watch?v=J1EuvLs3Fsg&t=10s	Video de César Rodríguez Cano, profesor de la facultad de Ciencias Políticas y Sociales UNAM, en el que muestra la definición de plagio, cómo buscar fuentes documentales, cómo citar y un ejemplo de cómo escribir una cita sin caer en plagio.
University of New South Wales https://www.student.unsw.edu.au/skills	Sitio en inglés que presenta estrategias para la elaboración de trabajos escritos de investigación que sirven de apoyo a estudiantes y profesores.

CONCLUSIONES Y REFLEXIONES FINALES

- La investigación es una estrategia que permite generar nuevas experiencias que contribuyen a la comprensión de un tema y comprobar el nivel de autorregulación que alcanzó el estudiante. Además de permitirles reforzar lo que aprendieron durante el curso y a fortalecer sus habilidades para analizar y mejorar su propio desempeño.
- La investigación debe contemplarse como la estrategia de evaluación principal, lo que implica que esté presente durante todo el curso y no solo como un producto final.
- Es importante considerar que una investigación requiere gran cantidad de tiempo tanto para los estudiantes como para el docente, por lo que se sugiere tomar en cuenta periodos adicionales en la elaboración del plan de trabajo.
- Promover la revisión de materiales entre pares en cada etapa de la investigación, para que los estudiantes reciban retroalimentación y solucionen los problemas que predominen.
- La retroalimentación efectiva y oportuna es vital, ya que permite al estudiante alcanzar los objetivos y las expectativas que se plantea con la investigación. Además de promover la evaluación del análisis, solución de problemas, pensamiento crítico, autoevaluación del proceso de aprendizaje, entre otros.

REFERENCIAS

Díaz-Barriga, F. y Hernández-Rojas, G. (2002). *Estrategias docentes para un aprendizaje significativo. Una interpretación constructivista*. Ciudad de México, México: McGraw-Hill.

Durante-Montiel, I., Lozano-Sánchez, J., Martínez-González, A., Morales-López, S. y Sánchez-Mendiola, M. (2012). Ejemplo para la evaluación de análisis crítico de artículos. En *Evaluación de competencias en ciencias de la salud.* Ciudad de México, México: Médica Panamericana.

Ko, S., y Rossen, S. (2017). Course design and development. En: Teaching online: *A practical guide* (pp. 33-91). Routledge.

Martínez-González, A. y Sánchez-Mendiola, M. (2015). La pregunta de investigación en educación médica. *Investigación en Educación Médica*, *4*(13), enero-marzo 2015, 42-49. http://riem.facmed.unam.mx/index.php/riem/article/view/324

Matos-de Rojas, Y. y Pasek de Pinto, E. (2008) La observación, discusión y demostración: técnicas de investigación en el Aula. *Laurus*, *4*(27), 33-52. http://www.redalyc.org/articulo.oa?id=76111892003

Peimani, N. y Kamalipour, H. (2021) Online Education and the COVID-19 Outbreak: A Case Study of Online Teaching during Lockdown. *Education Sciences*, *11*, 72. https://doi.org/10.3390/educsci11020072

Race, P. (2019). Designing assessment. Reports. En: *The Lecturer's Toolkit. A practical guide to assessment, learning and teaching* (89-92). Nueva York: Routledge.

Rodríguez, J. y Cordero, N. (2011). *Desarrollo de rúbricas para evaluar las investigaciones que hacen los estudiantes.* Oficina de Evaluación del Aprendizaje Estudiantil Decanato de Asuntos Académicos-UPR-RP. https://bit.ly/2CEz02T

Verdejo, P., Encinas, M. y Trigos, L. (2011) Estrategias para la evaluación de aprendizaje complejos y competencias. En *Estrategias para la evaluación de aprendizaje complejos y competencias* (pp. 19-45). https://bit.ly/33EyNbM

Capítulo 32

DIARIO DE CAMPO

Adrián Martínez González, Guadalupe Soto Estrada

"Aprende de ayer, vive para hoy, ten esperanza en el mañana.
Lo importante es no dejar de cuestionar.
La curiosidad tiene sus propios motivos para existir."
Albert Einstein

INTRODUCCIÓN

La formación académica implica no solamente el aprendizaje de conocimientos, que sin duda son esenciales, sino, además el desarrollo de habilidades y la conformación de actitudes y valores en los estudiantes, por lo que resulta indispensable contar con instrumentos, estrategias y técnicas que les permitan a los docentes obtener información acerca del progreso de los alumnos, y así tomar las mejores decisiones para alcanzar el logro de los aprendizajes esperados.

La utilización sistemática de instrumentos cualitativos, como es el diario de campo, todavía es poco frecuente en educación media superior en comparación con su empleo en la educación superior, particularmente en las áreas sociales y de la salud, como por ejemplo en Biología, Trabajo social, Psicología, Medicina, Antropología, entre otras. Dada la naturaleza descriptiva y reflexiva del diario de campo, este instrumento cumple paralelamente dos funciones pedagógicas: ser facilitador tanto de la enseñanza como del aprendizaje, y valorar elementos conceptuales, cognitivos, emocionales y éticos de los alumnos mediante la escritura del documento.

Este texto tiene como objetivo presentar de manera clara y sucinta información del diario de campo como instrumento de evaluación cualitativa, así como orientar y ejemplificar su uso.

El documento está organizado en ocho secciones. En la primera se explica qué es el diario de campo, en la segunda se detallan los pasos para diseñar una evaluación mediante este instrumento, en la tercera se recomienda cómo aplicarlo, en la cuarta se hacen algunas sugerencias para utilizar el diario de campo a distancia de forma electrónica, en la quinta se muestra cómo analizar sus resultados, en la sexta se dan algunos ejemplos, en la séptima se recomiendan algunos recursos en línea que pueden complementar la información presentada en este capítulo y en la octava se presentan las conclusiones y reflexiones finales.

Este texto es un recurso de autoestudio, reflexión y consulta que puede leerse o no en orden. Pretende ser un punto de partida para que el profesor explore las posibilidades que ofrece el diario de campo, reflexione sobre sus prácticas de evaluación e intercambie ideas con otros colegas.

¿QUÉ ES?

- El diario de campo es un instrumento del, para y cómo aprendizaje que permite al profesor conocer mediante escritos, registros descriptivos, analíticos y críticos los comportamientos, las actividades, los eventos y otras características que suceden durante una observación o una práctica con la finalidad de aplicar los conocimientos teóricos adquiridos. Además, contribuye a que los estudiantes desarrollen otras capacidades relacionadas con actitudes, valores e inclusive aspectos éticos.
- En él, los alumnos registran aquellos acontecimientos que observan sobre un tema o una situación de interés y que les resultan especialmente significativos. Por ello, la observación y la reflexión son los dos elementos fundamentales en la elaboración del diario de campo. Se considera que al menos hay cuatro procesos formativos involucrados en el diario de campo: la apropiación del conocimiento, la metacognición, la competencia escritural y el sentido crítico.
- Como instrumento de evaluación, el diario de campo es elaborado por los estudiantes bajo criterios establecidos por el profesor o un consenso de varios profesores, los cuales permiten verificar el logro de los aprendizajes esperados.
- Puede incluir notas, dibujos, bocetos, esquemas, croquis y fotografías. Además, es recomendable que integre comentarios y sugerencias escritas por el profesor. Se considera que el hecho de escribir en un diario contribuye a reforzar conocimientos a la vez que permite la organización mental de lo que se desea transmitir.
- De acuerdo con Spradley (1980, citado en Wolfinger, 2002) para obtener un escrito comprensivo con una descripción detallada de lo que pasa en un periodo particular de tiempo, es recomendable recabar información sistemática sobre los siguientes elementos:

    - *Espacio*: lugares físicos.
    - *Actores:* las personas involucradas.
    - *Actividad:* una serie de acciones que las personas realizan.
    - *Objeto:* las cosas físicas que están presentes en el lugar.
    - *Acto:* acciones particulares que las personas efectúan.
    - *Evento:* una serie de actividades relacionadas que las personas llevan a cabo.
    - *Tiempo:* la secuenciación que tiene lugar en el tiempo.
    - *Metas:* fin que las personas tratan de lograr.
    - *Sentimientos:* las emociones que sienten y expresan tanto las personas observadas como el escritor.

- Es posible elaborar un diario de campo integral o uno específico, según lo considere más apropiado y útil el profesor.
- El diario de campo promueve la reflexión mediante el cuestionamiento de la propia práctica y la teoría que la guía, lo que enriquece el conocimiento de los alumnos e incluso transforma su práctica. Así, el diario de campo contribuye al desarrollo de competencias profesionales y permite al estudiante la identificación de áreas de oportunidad que pueden limitar el logro de sus metas.
- Mediante el análisis de su contenido es posible entrever los conocimientos, las impresiones, las expectativas, las ideas, los valores, los deseos, las confusiones y las dificultades de su autor, lo que promueve la actividad reflexiva sobre el propio aprendizaje.
- Algunas fortalezas de la evaluación mediante el diario de campo incluyen: la posibilidad de plasmar evidencias sobre las reflexiones del alumno, lo que permite conocer el grado en que ha alcanzado los aprendizajes esperados; permite conocer las actitudes, inquietudes, dificultades e intereses que experimenta el alumno mediante su proceso de aprendizaje, por lo tanto, favorece que el estudiante se percate de sus propios procesos cognitivos y afectivos, y finalmente favorece el desarrollo de la expresión escrita y de ideas.
- Quizá una desventaja de la evaluación mediante el diario de campo es que, a pesar de que el empleo de otros instrumentos como la rúbrica o la lista de cotejo ayudan a sistematizar la información recabada en el diario de campo para emitir juicios de valor objetivos, es necesario dedicar un periodo considerable de tiempo para analizar y valorar las reflexiones que se plasmaron.

¿CÓMO LO DISEÑO?

Se sugiere que el profesor tome como eje rector en la planeación de sus clases y del ciclo escolar el mapa curricular, el perfil de egreso y el programa de su asignatura, módulo o curso; documentos en los que se estipulan los conocimientos, las habilidades, las actitudes y los valores que se pretende que el alumno adquiera y desarrolle. Con base en lo anterior, se podrá identificar la técnica, estrategia o instrumento más pertinente para evaluar el logro de los aprendizajes esperados.

Planeación

1) **Identifique los aprendizajes esperados que se pueden evaluar mediante el diario de campo. En lo posible diséñelo con otros profesores como cuerpo colegiado (alrededor de cinco).**
   ¿El diario de campo es el mejor instrumento para valorar esos aprendizajes?
2) **Decida cómo utilizará el diario de campo: evaluación formativa o sumativa o ambas.**
   De acuerdo con su programa de estudio, elija si el diario de campo se empleará para algunas actividades o como un cuaderno de trabajo de todo el curso. También es necesario determinar si el diario de campo será físico o electrónico, así como los tiempos de revisión y/o entrega.

3) **Determine los apartados que integrarán el diario de campo.**
Elija las secciones que le proporcionen información valiosa sobre el proceso de aprendizaje. Por ejemplo, descripciones, opiniones, impresiones y emociones, dudas, dilemas, toma de decisiones, formulación de hipótesis, argumentaciones, interpretaciones, entre otras. Una característica deseable en los alumnos es que adquieran sentido crítico por lo que es recomendable incluir siempre una sección de conclusiones en la que se evidencie la reflexión del autor.
4) **Defina los criterios con los que evaluará el diario de campo.**
Se recomienda que, desde el inicio del ciclo escolar, el alumno conozca de manera clara la forma en que será evaluado. Algunos criterios de evaluación pueden ser: redacción, ortografía, orden, exposición clara de ideas, precisión terminológica, emisión de puntos de vista y reflexión propia sobre el tema, entre otros.
5) **Elabore un documento de especificaciones para sus estudiantes.**
Redacte un documento en el que especifiquen los elementos que deberá contener el diario de campo, la forma en que se evaluará, las fechas de entrega y el valor de este instrumento en la calificación final. Se sugiere que en el diario de campo se contemplen al menos los siguientes elementos: Identificación (datos sobre el autor, la institución, la asignatura, el módulo, el curso, etcétera.), el reporte de actividades y una reflexión o conclusiones por parte del autor. También pueden incluirse instrucciones de como se encuentra estructurado o como se recomienda revisarlo.
6) **Seleccione el instrumento con el que evaluará el aprendizaje.**
Dependiendo de la información que recabe, determine el instrumento que le será de mayor utilidad para analizar el aprendizaje de sus alumnos, lo más sencillo y práctico puede ser una rúbrica o lista de cotejo.

Recomendaciones:

- Incluir siempre una sección de conclusiones en la que se evidencie la reflexión del autor.
- Se sugiere que desde el inicio de ciclo escolar el alumno conozca de manera clara la forma en que será evaluado.

Diseño

1) **Solicite a sus alumnos un cuaderno o carpeta en la que se asentarán sus registros, o bien, especifique el recurso virtual que se empleará** (ver más adelante en la sección de “El diario de campo en la evaluación a distancia”)**.**
Es importante que, si se trata de un diario de campo físico, el alumno tenga oportunidad de seleccionar el color, el tamaño y el diseño de su cuaderno o carpeta.
2) **Especifique las secciones que deben desarrollarse en el diario de campo, establezca la longitud máxima para desarrollar cada uno de los rubros.** Existen distintos tipos de diario, sin embargo, algunos de los más útiles como instrumentos de evaluación pueden ser los siguientes:

- *Diario científico o de investigación:* El objetivo principal es registrar observaciones y datos precisos relacionados con un proyecto de investigación.
- *Diario mixto:* Además de registrar observaciones, se pueden agregar las reflexiones del estudiante, así como impresiones, sensaciones o puntos de vista.
- *Diario de clase:* Puede ser elaborado tanto por los docentes como por los estudiantes para registrar de una forma sistemática las actividades, experiencias e inclusive problemas que se presentaron durante el proceso de enseñanza-aprendizaje-evaluación.
- *Diario de comunidad:* Se elabora considerando los objetivos y actividades a cumplir en escuelas, por ejemplo, sobre pláticas de educación para la salud a escolares o a padres de familia, levantamiento de croquis, participación en campañas de vacunación, entre otras.

3) **Explique el tipo y profundidad de información que espera encontrar en cada una de las secciones.** Ejemplifique el tipo de descripciones, interpretaciones y reflexiones que evaluará en el diario de campo.
4) **Delimite las fechas de entrega del diario de campo y realimente a los estudiantes.** Si se trata de un instrumento que servirá para la evaluación es importante señalar de manera oportuna (de preferencia al inicio del curso) las fechas de corte, revisión y entrega del diario de campo. Recuerde que, en un proceso formativo, la retroalimentación constante del profesor sobre los registros, observaciones, comentarios, dudas y reflexiones del alumno es fundamental para mejorar el aprendizaje.
5) **Considere un instrumento de autoevaluación, coevaluación o heteroevaluación.** Se recomienda considerar que, además de la evaluación por parte del docente, se fomente que los alumnos se evalúen a sí mismos y entre pares. La valoración que realice el alumno de su propio proceso de aprendizaje les permitirá detectar, a él y a usted, sus áreas de oportunidad, además de desarrollar una actitud autocrítica de sus fortalezas y debilidades que le ayudarán durante toda su vida.
6) **Programe sesiones de retroalimentación.** Destine periódicamente en su plan de trabajo cierto tiempo para indicar, al grupo y de forma individual, los temas con mayor facilidad y los de mayor complejidad, así como posibles actividades que contribuyan a esclarecer las dudas que vayan surgiendo a partir de la utilización del diario de campo.

Recomendación

- En algunas ocasiones, al iniciar la actividad de aprendizaje, realice una dinámica para que los alumnos tengan la posibilidad de expresar algunas de las ideas que plasmaron en su escrito.

Revisión

- Si es la primera vez que se utiliza el diario de campo de forma sistemática como instrumento para la evaluación del aprendizaje, se sugiere tomar notas sobre las mejoras que haría en la planeación, diseño e implementación de este para futuros cursos.

- Al finalizar el ciclo escolar, se debe preguntar a los estudiantes su opinión respecto al diario de campo como instrumento para dar seguimiento a su progreso en el aprendizaje, así como los obstáculos que enfrentaron para elaborarlo.
- Se sugiere compartir con sus colegas su experiencia respecto a este instrumento como práctica de evaluación formativa y/o sumativa.

¿CÓMO LO APLICO?

- Cuando el profesor determina que va a emplear el diario de campo debe explicar de forma clara a los alumnos la estructura que debe tener, las habilidades que desarrollarán mediante su elaboración y los criterios que se tomarán en cuenta para su valoración.
- El diario de campo es ideal para evaluaciones de tipo formativo, sumativo o ambas. De tipo formativo, permite entrever la evolución del alumno durante el ciclo escolar, identificar sus áreas de oportunidad y fortalezas, así como realimentar su proceso de aprendizaje; con el tipo sumativo, en el que se puede asignar un valor cuantitativo, con apoyo de otras técnicas, estrategias o instrumentos de evaluación, permite valorar el desempeño del alumno y plasmarlo en una calificación.
- Al ser un documento que exige un registro de acontecimientos–sucesos–observaciones, promueve que el alumno reflexione por medio de sus opiniones, creencias, dudas, sentimientos, concepciones y demás pensamientos y emociones del autor. Se sugiere que se elabore de forma individual.
- En su implementación, se recomienda que el profesor ejemplifique en las primeras sesiones un diario de campo que él elabore. De esta forma, puede comenzar las primeras clases leyendo la descripción de su experiencia personal, la forma en que podría explicar la situación o fenómeno de estudio, las teorías o información a las que debe remitirse para entender el fenómeno, los sentimientos que experimentó, las dudas que le surgieron, entre otros aspectos. Así, tendrá la posibilidad de mostrar a los alumnos los elementos que espera que contenga el diario de campo.
- Se sugiere que, mediante el empleo del diario de campo, el profesor contraste las concepciones teóricas del estudiante al inicio del proceso educativo y analice como éstas se modifican o enriquecen a medida que avanza y finaliza el curso, lo que contribuye a conocer de forma más precisa el progreso y las necesidades de cada alumno con el objetivo de potenciar su aprendizaje.

EL DIARIO DE CAMPO EN LA EVALUACIÓN A DISTANCIA

En la actualidad y ante el gran reto que implica trasladar los espacios educativos a espacios virtuales en los que la evaluación debe ser de calidad y congruente con los objetivos de aprendizaje, el diario de campo representa una gran herramienta para valorar el desarrollo del alumno y sirve al docente para dar retroalimentación efectiva que promueva la mejora continua en los aprendizajes y habilidades adquiridas.

¿Cómo diseño el diario de campo a distancia?

Aunque el diario de campo tradicional se realiza en papel, el empleo de diversos recursos tecnológicos permite diversificar sus elementos y la forma en que se plasman, ya que, además de hacer registros escritos, se pueden emplear otro tipo de recursos como la grabación de notas de voz o incluso la grabación de videos, mismos que pueden ser muy útiles para la evaluación formativa y sumativa.

Se recomienda que el docente especifique los apartados que debe contener el diario de campo de acuerdo con los objetivos de aprendizaje y el contexto en el cual se desarrollará.

De ser posible explique a los estudiantes lo que espera que ellos plasmen en el diario de campo y ejemplifique el tipo de descripciones, interpretaciones y reflexiones que deben incluir.

Establezca con anticipación las fechas de elaboración, revisión y retroalimentación del diario de campo.

Defina los criterios, así como el tipo de evaluación que se realizará (autoevaluación, coevaluación o heteroevaluación).

Para entender un poco más sobre el diseño del diario de campo le recomendamos que revise el siguiente tutorial (https://www.youtube.com/watch?v=NDa7AdAGdvM) en el que se explican los elementos que lo conforman.

¿Cómo instrumento el diario de campo a distancia?

Se puede partir de un formato básico con el apoyo de algún software conocido por la mayoría de los estudiantes, por ejemplo, el de Microsoft Word, sin embargo, actualmente se encuentran disponibles una gran variedad de recursos para mejorar la experiencia de los estudiantes al elaborar un diario de campo. En el siguiente enlace se puede descargar un formato de diario de campo que servirá como ejemplo para la utilización con sus estudiantes: Formato de diario de campo. (https://www.academia.edu/9405658/FORMATO_DE_DIARIO_DE_CAMPO_DIARIO_DE_CAMPO).

Se recomienda elegir la opción más adecuada de acuerdo con las necesidades y objetivos del diario de campo solicitado, además, es importante considerar el equipo de cómputo con el que trabajarán los alumnos (computadora personal, laptop, tableta, teléfono celular, etc.), así como la plataforma elegida, debido a que, aunque la mayoría de estos recursos son multiplataforma, algunos tienen ciertas limitaciones.

Se sugiere que exista cierta apertura para que los estudiantes, de acuerdo con sus recursos y necesidades, elijan entre algunas plataformas propuestas por el docente.

También es importante considerar que, en muchos casos, es suficiente la versión gratuita de las plataformas para elaborar adecuadamente el diario de campo, lo que facilita el acceso por parte de los alumnos.

Al utilizar formatos en línea para elaborar un diario de campo, es importante no perder de vista el propósito que se persigue.

A continuación, se describen algunas de las plataformas más utilizadas en las que se puede elaborar un diario de campo virtual:

Google Keep

Para utilizar este recurso es necesario tener una cuenta de Google y en esta se pueden crear notas a las que se puede asignar diferentes parámetros, etiquetas o recordatorios. Una de las principales ventajas es que puede añadir directamente una nota de voz o una imagen, se puede archivar y compartir con otros usuarios. Dentro de sus principales funciones están la de crear listas con casillas de verificación, crear recordatorios, compartir las notas elaboradas e inclusive crear notas colaborativas para editar entre varios usuarios.

También permite destacar con un color diferente cada nota y organizarlas al asignarles etiquetas. Otra ventaja es que permite crear notas a mano, escribiendo directamente sobre la pantalla de un dispositivo móvil.

Simplenote

Este recurso permite tomar notas de texto sencillas que se encuentran sincronizadas entre todos los dispositivos. Es una aplicación multiplataforma, que permite etiquetar las notas para ubicarlas más fácilmente. Además, permite la edición simultánea entre varios usuarios e incluso publicar en Internet a través de la generación de una URL. Aunque una de las desventajas es que no permite organizar notas mediante carpetas.

Microsoft OneNote

Es una aplicación web de Office que resulta útil para tomar apuntes, hacer listas y hacer anotaciones de prácticamente cualquier cosa. Se descarga de forma gratuita en cualquier dispositivo y permite organizar notas por tema a través de la creación de blocs de notas, que son como libretas o cuadernos compuestos de una o varias notas, los cuales se pueden agrupar por secciones. Se trata de un recurso fácil de usar que permite agregar imágenes, vínculos, listas y etiquetas.

Evernote

Se trata de una herramienta web que cuenta con una versión gratuita y una de pago que permite almacenar y compartir información a la que se puede acceder desde una computadora o dispositivo móvil con acceso a Internet, además, permite guardar todo tipo de información y organizarla de forma sencilla porque se puede clasificar y etiquetar a las notas. Y al igual que otras de las herramientas ya descritas, también se puede añadir texto (que puede incluir dictados, ideas, manuscritos, enlaces.) y otro tipo de elementos como notas de voz o imágenes.

Otras ventajas de este recurso son que el sistema de reconocimiento de texto permite identificar las palabras que contiene una imagen capturada de algún documento y la información se puede organizar mediante libretas y etiquetas.

Finalmente, su sistema de búsqueda resulta bastante útil ya que, aunque las notas se pueden encontrar con el nombre de las etiquetas o libretas, también se pueden localizar por las palabras que se encuentran al interior de las notas.

¿CÓMO ANALIZO SUS RESULTADOS?

- El profesor puede usar otros instrumentos, técnicas o estrategias para valorar el diario de campo, como: la lista de cotejo, la rúbrica, la guía de observación, entre otros, los cuales aportan información valiosa para realimentar el aprendizaje del estudiante y del grupo.
- A nivel personal, el profesor puede identificar el avance de cada alumno y realimentar sobre los conocimientos y habilidades que ha desarrollado favorablemente, así como los temas en los que tiene dificultades.
- A nivel grupal, el profesor puede tener un panorama sobre el grado de avance que ha tenido el grupo, si se han alcanzado aprendizajes esperados de las unidades, así como las áreas de mejora y el establecimiento de acciones remediales o para fortalecer el aprendizaje.

EJEMPLOS

Tabla 1. Ejemplo de lista de cotejo para un diario de campo

Instrucciones: La siguiente lista de cotejo presenta los elementos que deben contener las diferentes secciones del diario de campo. Es útil para profesores y estudiantes. Marque con una ✔ el recuadro del elemento que se incluyó en el diario de campo.

Portada	
☐	Institución o Universidad.
☐	Facultad.
☐	Asignatura.
☐	Nombre del proyecto.
☐	Nombre del autor.
☐	Año de elaboración.

Petición de devolución del diario de campo (Si se encuentra en físico)	
☐	Solicitud para devolver el diario de campo.
☐	Dirección en la que se tiene que entregar.
☐	Teléfono/correo electrónico para contactar a quien extravió el diario de campo.
☐	Agradecimiento por la devolución del diario de campo.
Reportes de actividades	
☐	Fecha (día, mes, año) de la observación.
☐	Objetivo de la observación.
☐	Registro de información descriptiva.
☐	Registro de información reflexiva.

Fuente: Elaboración propia

Tabla 2. Ejemplo de lista de cotejo para autoevaluar un diario de campo

Instrucciones: A continuación, se presentan los aspectos que debes tomar en cuenta durante la elaboración de tu diario de campo en la información descriptiva y reflexiva. Marca con una ✔ los elementos que contemplaste en el desarrollo de tu diario de campo. En caso de que te hayan faltado algunos elementos puedes poner las razones en la parte de observaciones.

Información descriptiva	☐	Registré la fecha en que llevé a cabo la observación.
	☐	Anoté el lugar donde realicé la observación.
	☐	Describí el escenario físico donde llevé a cabo la observación.
	☐	Describí el entorno social y la manera en que las personas interactuaron dentro del escenario observado. Esto incluye la comunicación no verbal y patrones específicos de conducta como conflictos, toma de decisiones o colaboración entre personas, etcétera.
	☐	Describí a las personas y los roles de cada uno en el escenario.
	☐	Describí el significado de lo que observé desde la perspectiva de las personas.
	☐	Registré frases exactas o aproximaciones muy cercanas de los comentarios que hicieron las personas y que tienen relación directa con mi objetivo del estudio.
	☐	Describí cualquier impresión o emoción que me causó la situación que observé.
Información reflexiva	☐	Incluí mis ideas, impresiones, pensamientos y críticas respecto de lo que observé.
	☐	Incorporé los cuestionamientos o inquietudes que me surgieron a raíz del análisis de lo que observé.
	☐	Esclarecí dudas, errores o malentendidos que tenía en mis notas.
	☐	Añadí lo que observé y especulé, así como las razones por las que concluí que ese fenómeno ocurrió.
	☐	Registré los pensamientos que he estado considerando para futuras observaciones.
Observaciones:		

Fuente: Elaboración propia

Tabla 3. Ejemplo de rúbrica para evaluar la información reflexiva de un diario de campo

Instrumento para el profesor

Instrucciones: La siguiente rúbrica muestra cuatro categorías de desempeño –deficiente, regular, bueno y excelente– respecto al contenido que plasma el estudiante en su diario de campo. Marque con una ✔ la categoría que corresponda, de acuerdo con el desempeño que mejor caracteriza el trabajo de su alumno.

Aspectos a evaluar en el diario de campo	Excelente (3)	Bueno (2)	Regular (1)	Deficiente (0)
Opiniones personales	Escribe abundantes opiniones personales.	Escribe suficientes opiniones personales aunque no en todos los registros.	Escribe algunas opiniones personales.	Escribe pocos comentarios y opiniones personales.
Juicios valorativos	Emite bastantes juicios valorativos de las situaciones e interacciones que observó y analizó.	Presenta algunos juicios valorativos de las observaciones que realizó y analizó.	Presenta pocos juicios valorativos de las observaciones que realizó.	No presenta ningún juicio valorativo de las observaciones que realizó.
Descripción de decisiones tomadas	Describe todas las decisiones que tomó conforme realizó su diario de campo.	Describe algunas de las decisiones que tomó conforme realizó su diario de campo.	Presenta pocas decisiones que tomó conforme realizó su diario de campo.	No presenta ninguna decisión que tomó conforme realizó su diario de campo.
Formulación de hipótesis	Formula hipótesis e identifica teorías que le permiten explicar el fenómeno observado.	Formula una hipótesis del fenómeno observado, pero no identifica ninguna teoría que le permita explicarlo.	Presenta un atisbo de una posible hipótesis del fenómeno observado, pero no logra identificar ninguna teoría que le permita explicarlo.	No formula ninguna hipótesis ni presenta ninguna teoría que le permita explicar el fenómeno observado.
Descripción de impresiones y emociones	Plasma suficientes impresiones y emociones que le generaron las observaciones que realizó.	Plasma algunas de las impresiones y emociones que le generaron las observaciones que realizó.	Las impresiones y emociones que le generaron las observaciones que llevó a cabo son muy escasas.	No plasma ninguna impresión ni emoción generada a partir de las observaciones que realizó.
Dudas y dilemas	Plantea en todos sus registros las dudas y dilemas a los que se enfrentó.	Plantea en algunos de sus registros las dudas y dilemas a los que se enfrentó.	Plantea pocas dudas y dilemas que le surgieron.	Carece del planteamiento de dudas y dilemas a los que se enfrentó.
Observaciones:				
Puntaje máximo: 18 **Puntaje mínimo: 0**				

Fuente: Elaboración propia

RECURSOS EN LÍNEA

Tabla 4.

Recurso	Descripción
Rubric for field notes/field books https://serc.carleton.edu/NAGTWorkshops/field/assessments/48524.html	Rúbrica para evaluar el trabajo de los alumnos mediante un diario de campo. En ella se presentan expectativas claras de lo que se espera de los alumnos antes de que ellos inicien su diario de campo. Permite ubicar el desempeño de los alumnos en diferentes niveles de desempeño -puede mejorar, satisfactorio y ejemplar- respecto a su conocimiento, habilidades y actitudes.
Field Notes Rubric http://csmgeo.csm.jmu.edu/geollab/Whitmeyer/IrelandDocuments/ExampleGradingRubrics.pdf	Este documento presenta una pequeña compilación de ejemplos de rúbricas del campo de la Geografía.
Field Study Rubric https://www.livetext.com/doc/3025353	Es una rúbrica de autoevaluación en el que se valora la introducción y las reflexiones finales, las entrevistas, el trabajo de investigación y la síntesis que realizó el alumno en la elaboración de su diario de campo. Se presentan cinco gradientes de desempeño.
Videos	
Tema: diario de campo https://youtu.be/gmJ89cHP42g	En este video se describen las principales características y secciones, así como la utilidad de un diario de campo.
Diario de campo https://youtu.be/XqtAiqLXoDs	En este video se mencionan las principales características del diario de campo y da algunas recomendaciones de cómo realizarlo.
Ejemplo y Descripción Diario de Campo https://youtu.be/DXVoHYuP2Qo	En este video se describen las características, pasos y recomendaciones para hacer un diario de campo en formato electrónico.

CONCLUSIONES Y REFLEXIONES FINALES

- Al utilizar el diario de campo, es importante especificar los elementos que debe contener y establecer criterios de evaluación claros y precisos; aunque sin perder flexibilidad, para proporcionar al alumno cierta autonomía en el diseño y elaboración de este recurso.
- El empleo del este recurso para el proceso de enseñanza-aprendizaje-evaluación favorece el desarrollo y el fortalecimiento de habilidades de metacognición y aporta información útil para realimentar a los alumnos sobre su capacidad de escritura, análisis y síntesis.
- El diario de campo sirve como un instrumento de evaluación, heteroevaluación y autoevaluación porque le permite al docente contar con información útil para monitorear el avance del alumno y brindar retroalimentación, a los alumnos les da oportunidad de evaluar el desempeño de sus pares, además de monitorear su propia evolución de aprendizaje y reflexionar sobre aspectos que el profesor le señala mediante preguntas, comentarios u observaciones.

REFERENCIAS

Albertín-Carbó, P. (2007). La formación reflexiva como competencia profesional. Condiciones psicosociales para una práctica reflexiva. El diario de campo como herramienta. *Revista de Enseñanza Universitaria,* 30, 7-18. Recuperado de http://institucional.us.es/revistas/universitaria/30/Albertin.pdf

Alzate-Yepes, T., Puerta-Catano, A. y Morales, R. (2008). Una mediación pedagógica en educación superior en salud. El diario de campo. *Revista Iberoamericana de Educación, 47*, 1-10. Recuperado de https://rieoei.org/historico/deloslectores/2541Alzate.pdf

Alzate-Yepes, T. (s. f.). El diario de campo como mediación pedagógica en educación superior. Universidad de Valencia-España. Recuperado de https://giiesen.files.wordpress.com/2010/04/art1-diario_de_campo_pdpd.pdf

Durante-Montiel, M., Lozano-Sánchez, J., Martínez-González, A., G., Morales-López, S. y Sánchez-Melchor, M. (2012). *Evaluación de competencias en ciencias de la salud.* Ciudad de México: Editorial Médica Panamericana.

Martínez-Rodríguez, L. A. (2007). La observación y el diario de campo en la definición de un tema de investigación. *Perfiles libertadores*, 73-80. Recuperado de https://www.ugel01.gob.pe/wp-content/uploads/2019/01/1-La-Observaci%C3%B3n-y-el-Diario-de-campo-07-01-19.pdf

Valverde-Obando, L. A. (1993). El diario de campo. *Revista Trabajo Social*, 39, 308-319. Recuperado de http://www.binasss.sa.cr/revistas/ts/v18n391993/art1.pdf

Wolfinger, N. H. (2002). On writing fieldnotes: collection strategies and background expectancies. *Qualitative Research, 2*(1), 85-95. Recuperado de https://entwicklungspolitik.uni-hohenheim.de/uploads/media/Day_2_-_Reading_text_4_02.pdf

Fuentes de información utilizadas en la descripción de los recursos en línea para elaborar un diario de campo:

Freire D. (2021) Google Keep: qué es y para qué sirve. (Internet) [Consultado el 19 de septiembre de 2021]. Recuperado de https://andro4all.com/guias/google/google-keep-que-es-y-para-que-sirve

Google Keep; Guarda tus ideas estés donde estés. (Internet) [Consultado el 20 de septiembre de 2021]. Recuperado de https://www.google.com.mx/keep/

Carbonell L. (2018). Simplenote o como tomar notas en cualquier dispositivo. (Internet) [Consultado el 20 de septiembre de 2021]. Recuperado de https://atareao.es/software/productividad/simplenote-o-como-tomar-notas/

Toledo A. (2019). Simplenote. (Internet) [Consultado el 22 de septiembre de 2021]. Recuperado de https://simplenote.uptodown.com/windows

Sören RA. (2017). El diario de campo en la era digital (Evernote). (Internet) [Consultado el 20 de septiembre de 2021]. Recuperado de https://isdfundacion.org/2017/11/09/etnografia-virtual-no-3-el-diario-de-campo-en-la-era-digital-evernote/

Microsoft OneNote. (2021). (Internet) [Consultado el 18 de septiembre de 2021]. Recuperado de https://www.microsoft.com/es-mx/microsoft-365/onenote/digital-note-taking-app

Capítulo 33

DEBATE

Elibidú Ortega Sánchez, Enrique Ricardo Buzo Casanova

"El objetivo de la argumentación o de la discusión,
no debe ser la victoria sino el progreso."
JOSEPH JOUBERT

INTRODUCCIÓN

A lo largo de la historia de la humanidad, los debates han sido la forma predilecta de intercambio de puntos de vista en ámbitos formales, académicos o políticos. En ellos las personas acuden a la palabra para exponer sus opiniones, objeciones o puntos de vista, con la esperanza de convencer a los demás, en lugar de recurrir a la violencia para imponerse sobre ellos. De hecho, para garantizar que el debate sea organizado, debe haber un moderador que asigne turnos y vele por el mutuo entendimiento (Raffino, M., 2021).

En el ambiente escolar, el debate es una técnica de evaluación cualitativa que permite observar las capacidades del alumno para argumentar en una exposición como pueden ser contenidos, capacidad de argumentación, claridad y precisión conceptual; al igual que ciertas actitudes como respeto, tolerancia, capacidad para esperar su turno, cooperación, etcétera. Para debatir con éxito, es necesario combinar el estudio profundo de la materia sobre la cual se argumentará, con las capacidades de oratoria para exponer de manera convincente los propios argumentos, es por ello que el debate forma parte de las técnicas y métodos de enseñanza en la escuela y otras instituciones académicas. En el salón de clases, el debate se puede utilizar como una estrategia a trabajar con los estudiantes dentro del aula de clases, o se puede llevar al grupo a observar y escuchar debates relacionados con la profesión.

El propósito de este capítulo es desarrollar las habilidades de los profesores en el diseño, aplicación y análisis del debate como una herramienta de carácter cualitativo para realizar evaluaciones del, para y como aprendizaje más precisas a distancia o de manera presencial.

El capítulo está organizado en ocho secciones. En la primera se explica qué es el debate, en la segunda se abordan los pasos para su correcta planeación y diseño, en la tercera se detalla la forma de aplicación, en la cuarta se presenta cómo implementarlo a distancia, en la quinta se ve cómo analizar sus resultados, en la sexta se brindan dos ejemplos para aclarar

su uso, en la séptima se presentan algunos recursos en línea y en la octava tenemos las conclusiones y reflexiones finales.

¿QUÉ ES?

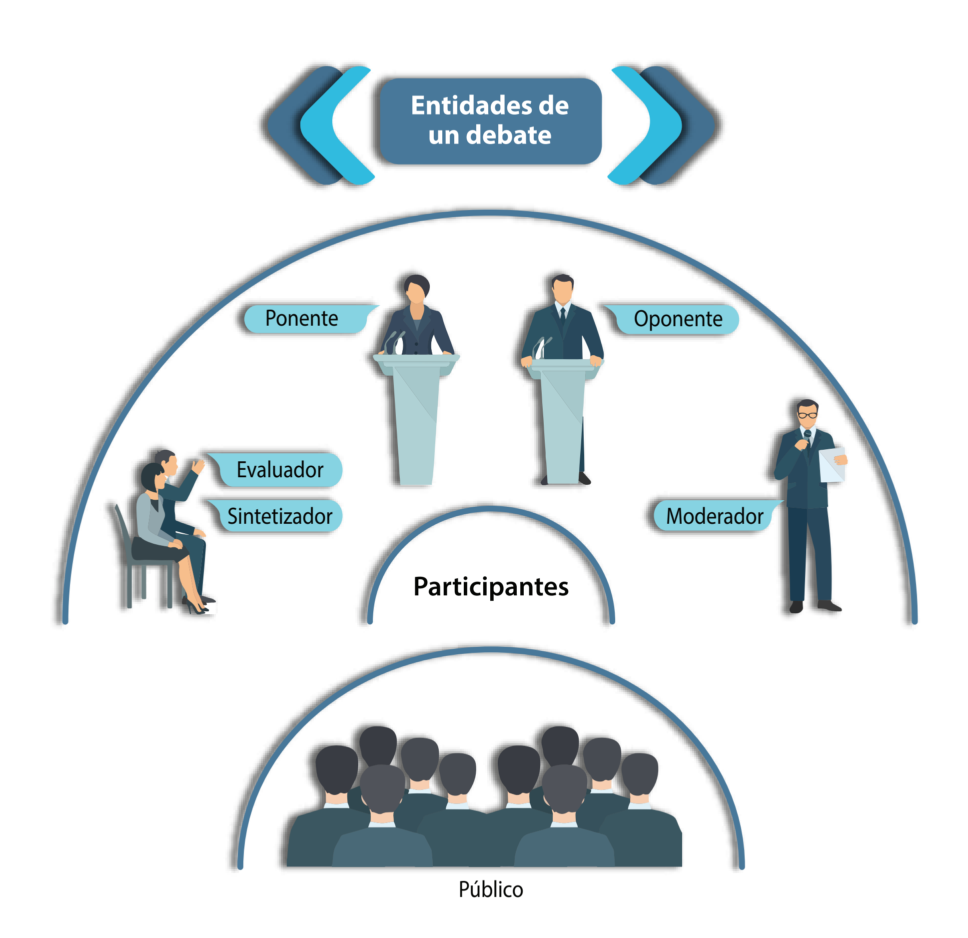
El debate es una forma de discusión formal y organizada que se caracteriza por el intercambio argumentado de ideas y/o puntos de vista entre dos o más personas con posiciones opuestas sobre un tema determinado (Centro de escritura académica y pensamiento crítico, 2015). Como su objetivo es enfrentar dos o más opiniones acerca de un determinado tema polémico, o al menos discutible desde diversos puntos de vista, precisa de una investigación documental rigurosa para poder replicar con fundamentos. Durante el debate, cada postura debe exponer su tesis y sustentarla por medio de argumentos y contraargumentos sólidos y claros con la dirección de un moderador para mantener el respeto y la objetividad entre ambas posturas. Además, cada posición debe buscar el interés del público, buscando que forme su opinión y contribuya en las conclusiones del mismo. Al término de este se deberá llegar a una conclusión a partir de la exposición organizada y pacífica de los argumentos.

El debate se integra por las siguientes entidades:

- **Participantes** (proponente y oponente): Pueden ser uno o varios, deberán limitarse al tema establecido y deben estar preparados para respaldar la postura que defiendan o para refutar los argumentos del individuo o grupo contrario. La calidad y fluidez del debate dependen de la capacidad de escuchar y contraargumentar, así como del conocimiento de los participantes.
- **Moderador:** Es indispensable para llevar a cabo un buen debate, ya que en él recae la responsabilidad de dirigir la discusión y que ésta se desarrolle de acuerdo con las reglas previamente establecidas y aceptadas por los oponentes; distribuir el tiempo de habla de manera equitativa y asignar los turnos, manteniendo el orden en el debate. También es responsable de dar inicio y concluir el debate. Por ello, es importante que el moderador también conozca sobre el tema, tenga capacidad de análisis y mantenga imparcialidad y tolerancia.
- **Público:** Es importante porque sus reacciones orientan a los participantes en sus argumentos; es decir, si el público reacciona de forma favorable, éstos saben que sus argumentos van en la dirección correcta o viceversa. También depende del público, tanto la orientación y enfoque del debate como el tipo de argumentos y lenguaje que se utilizará. Dependiendo de quién de los participantes haya sido más elocuente en sus argumentos, el público puede elegir un “ganador”.

De manera opcional pueden considerarse también:

- **Evaluador:** Debe observar con atención todos los aspectos del desarrollo del debate. Debe determinar la objetividad y responsabilidad con las que se han emitido las opiniones; analizar si se da una igualdad de oportunidades a los distintos sectores participantes del debate y si ha habido un tratamiento adecuado al tema.
- **Sintetizador:** Debe estar atento al desarrollo global del debate realizado. Debe, además, extraer sus propias conclusiones y a través de su análisis, dar cuenta de los aspectos deficitarios y las fortalezas, tanto en la organización del debate, como en la elaboración de argumentos esgrimidos por cada grupo. Por último, debe comunicar a los participantes el análisis realizado y las conclusiones finales.



Los debates son generalmente orales, aunque pueden darse también por escrito a través de plataformas de comunicación en línea apropiadas para ello, como son: *Stack Exchange, Loomio, Disqus, Moodle,* etcétera.

Dependiendo de si existe un conjunto oficial de reglas para debatir o si éstas se fijan de manera espontánea e improvisada, podemos distinguir entre los debates formales y los debates informales, respectivamente.

- **Debates formales:** Cuentan con reglas claras y preestablecidas, y un moderador que vela por el cumplimiento de éstas. En ellos, las formas deben cuidarse y generalmente el tema a discutir está muy bien acotado (Raffino, M., 2021).
- **Debates informales:** Caracterizados por la libertad argumentativa, no suelen estar acordados de antemano, ni contar con reglas formalmente establecidas. Tampoco suelen tener un moderador (Raffino, M., 2021).

Ventajas:

- Permite observar las capacidades del alumno para argumentar sobre el tema a discutir.
- Permite observar la capacidad de atención del grupo.
- Es útil para trabajar sobre las actitudes de respeto, tolerancia, capacidad para esperar su turno, cooperación, etcétera.

Limitaciones:

- Se necesita la atención total por parte del profesor o facilitador.
- Sin la atención necesaria, el grupo se puede salir de control.

A continuación, le explicaremos las consideraciones a fin de implementar el debate en una evaluación del y para el aprendizaje de manera presencial y a distancia.

¿CÓMO LO DISEÑO?

Para implementar el debate para la evaluación se recomienda analizar el programa de su asignatura para establecer los elementos que evaluará. Después se sugiere seguir el siguiente proceso:

Planeación

Se sugieren los siguientes puntos para planear el uso del debate como instrumento de evaluación.

1) **Determine los aprendizajes esperados que evaluará.**
   De acuerdo con el programa de estudios, curso o taller que esté impartiendo, especifique el contenido, habilidades, aptitudes, actitudes que valorará a través del debate.
2) **Seleccione el tipo de evaluación que desea realizar con el debate.**
   De acuerdo con el propósito de la evaluación puede ser diagnóstica, formativa o sumativa.
3) **Contemple a los participantes en el debate.**
   Considere si va a ser entre dos personas o entre dos grupos de personas.

4) **Determine el tipo de debate.**
Esto es fundamental para determinar la estructura, las normas o reglas que lo van a regir. Para identificar formatos de debate, se sugiere ver el video "Formatos de debate" https://www.youtube.com/watch?v=tF50jbdYAXI
5) **Defina el tema del debate.**
El tema debe estar íntimamente ligado a lo que se desea evaluar, debe ser un tema de interés que genere controversia.
6) **Identifique los recursos materiales que se van a necesitar.**
Prevea la infraestructura que requerirá ya sea que el debate se realice de manera presencial o a distancia.

Consejo 1. Pueden utilizar material de apoyo como imágenes, textos impresos o en Power Point, acetatos, etcétera.

7) **Plantee criterios claros de evaluación en un instrumento.**
Para que la evaluación del debate se lleve a cabo con sistematicidad, se recomienda el uso de otros instrumentos como la lista de cotejo o la rúbrica.

Consejo 2. Su elaboración debe realizarse con base en los aprendizajes esperados y debe especificar los criterios a evaluar con el debate.

Diseño

Con base en el tipo de debate que eligió durante la planeación, es necesario:

1) **Seleccionar el formato del debate.**
Decida si se permitirá la contraargumentación del equipo contrario de manera inmediata o después de la defensa de los planteamientos iniciales de los principales argumentos.

Consejo 3. Defina si se aceptarán preguntas del público presente y si éste recogerá de ellos apreciaciones del manejo de cada uno de los equipos en cuanto a los argumentos que son defendidos.

2) **Plantear las reglas del trabajo.**
Los tiempos, entregas, comportamiento durante el debate y los roles que asumirán los estudiantes.
3) **Preparar los contenidos teóricos de la argumentación.**
Solicite a los estudiantes que se informen del tema a través de textos, entrevistas, y otros medios de información confiables, con el fin de preparar la argumentación a ser utilizada en la presentación. Le recomendamos ver el video "Debate: Estructura y organización de las ideas" que se encuentra disponible en https://www.youtube.com/watch?v=USqZZ2QbT5Y

4) **Nombrar a los participantes.**
Además de las personas o grupos que van a debatir, nombre un moderador encargado de determinar el esquema de trabajo y a un secretario, si fuera necesario.
5) **Formar grupos a favor y en contra de los planteamientos.**
Defina los roles por grupo a partir del tema elegido, divida el curso en dos grupos, donde uno de ellos es el equipo defensor de una postura, y el otro constituye la contraparte.

Consejo 4. Si se desea asegurar la participación de todos los alumnos, es posible dividir el curso en grupos más pequeños, con el propósito de promover la argumentación y contraargumentación de los estudiantes.

Considere

- El nivel de estudio y características de sus alumnos para ajustar las características del debate en fondo y forma.
- Es importante asegurarse de que el tema del debate sea controversial, pero sin tocar temas sensibles para sus alumnos como puede ser lo relacionado con divorcio o muerte.
- Que debe tener muy claro el objetivo de la evaluación ya que el debate es un instrumento tan completo que puede fácilmente perder el foco de la misma durante su planeación y diseño.

Revisión

- Un ejercicio para enriquecer el debate es compartir el formato con otros profesores para recibir retroalimentación para mejorar este instrumento.
- Después de la aplicación, es muy útil que pida a los alumnos comentarios o sugerencias sobre el escenario, el tiempo de ejecución y sus aprendizajes alcanzados, entre otros aspectos.

¿CÓMO LO APLICO?

Existen diferentes maneras de llevar a cabo la aplicación de este instrumento de evaluación dependiendo de la cantidad de personas involucradas.

- Si decidió aplicarlo en parejas es necesario asignarles un tema para que lo discutan juntos. Posteriormente, ya ante el total del grupo, se le pide a un alumno que argumente sobre el tema a discutir, después de que el profesor lo indique debe continuar su compañero. El resto de los estudiantes deben escuchar con atención y tomar notas para poder debatir sobre el contenido.
- Si la aplicación será en grupo hay que dividirlo en dos partes; al azar pedirle a un equipo que busque argumentos para defender el contenido del tema y al otro equipo solicitarle que esté en contra. Después de un tiempo, cada equipo debe tratar de convencer al otro

de lo positivo de su postura con argumentos objetivos, ejemplos, dejando hablar a los otros, respetando los puntos de vista contrarios.

El profesor guiará la discusión y observará el comportamiento de los alumnos, anotando durante el proceso aspectos que le hayan llamado la atención y que le permitan realizar posteriormente una observación más dirigida.

En la aplicación del debate se distinguen las siguientes etapas (Raffino, M., 2021):

1) **Apertura.** A cargo del moderador, quien introduce el tema haciendo especial énfasis en su interés y actualidad y también las dos posturas que se enfrentarán en el debate. Luego hace lo mismo con los participantes, explicando su nivel de experticia en el asunto y su recorrido profesional o académico. También explica la dinámica a seguir y recuerda a los participantes las reglas, previamente establecidas.
2) **Cuerpo del debate.** Esta es la etapa que se asigna a la discusión del tema y cuyo protagonismo corresponde a quienes debaten. Ellos dispondrán generalmente de dos bloques de tiempo de igual extensión, para primero exponer un punto de vista, luego el otro, y finalmente exponer los argumentos y contraargumentos. La interacción entre los antagonistas se da en la sesión de preguntas y respuestas. Esto ayuda a mantener el orden y fomenta la capacidad de escuchar, tanto del público como de los participantes.

    - **Sesión de preguntas y respuestas.** Una vez expuesto el grueso del debate, el moderador, el público o los mismos participantes formulan algunas preguntas, de modo que los participantes puedan responderlas a su manera, pero siempre de forma ordenada. La sesión de preguntas y respuestas es una parte fundamental dentro del debate porque es en ella donde el público termina por inclinarse a favor o en contra de una postura. Además, es la oportunidad de los participantes para clarificar los puntos principales de sus argumentos, así como reforzar los puntos débiles. Las preguntas pueden estar a cargo del moderador, del público o de los mismos participantes.

    Consejo 5. La interacción entre los proponentes y oponentes debe ser mediada por el moderador cuando haga falta.

3) **Cierre del debate**

- **Consenso sobre las conclusiones.**
  En esta última etapa, el sintetizador hará un breve resumen de las posturas y comunicará la conclusión alcanzada en conjunto por los participantes.
- **Anuncio del ganador.**
  El evaluador dará el ganador o la posición dominante del debate.

Consejo 6. Cuando no existen las entidades del sintetizador y del evaluador el monitor estará a cargo de cerrar el debate al dar las conclusiones y el anuncio del ganador si es el caso.

EL DEBATE EN LA EVALUACIÓN A DISTANCIA

¿Será Internet un buen medio para debatir? En este apartado le mostramos como implementar el debate en un contexto en línea.

¿Cómo lo diseño a distancia?

Para el diseño del debate en la evaluación a distancia la planeación mantiene un proceso similar, sin embargo, deberá considerar la plataforma que mejor se ajuste al propósito del debate, así como a su formato.

A continuación, le vamos a dar una orientación general sobre cómo diseñar un debate con la aplicación *Loomio.*

1) Primero es necesario registrarse en la plataforma para tener una prueba gratuita de 7 días. Si se desea continuar tiene un costo que no es muy alto y es de por vida. Para el registro, vaya al sitio oficial: https://www.loomio.com/
2) Después de clic en **Try Loomio for free.**
3) Ingrese un correo electrónico porque le van a enviar un código de 6 dígitos a esa cuenta.
4) Ingrese el código, escriba su nombre y suba una foto.
5) Es necesario crear un nuevo grupo. Para ello complete los campos de nombre del grupo, gestionar y la categoría que describe mejor al grupo. De clic en **Crear grupo.**

¿Cómo lo instrumento?

Vamos a continuar con el ejemplo de la plataforma *Loomio*.

1) En la página principal, en la sección primeros pasos encontrará un botón que se llama **Comience un hilo de discusión.**

   Consejo 7. Los hilos son para discutir temas y obtener resultados claros.

Para crear el hilo es necesario completar la siguiente información y dar clic en **Nuevo hilo**.

- Título: Le sugerimos que sea el tema del debate.
- Notificar: Personas que participarán en el debate.
- Etiqueta: Corresponde a la categoría y se puede crear en el momento.
- Contexto: Descripción del debate.

2) Los debates más productivos comienzan con preguntas, así es que le sugerimos iniciar con la pregunta que detonará el debate.
3) Inicie un nuevo hilo para cada nueva consulta; una consulta puede incluir varias preguntas relacionadas. Utilice preguntas abiertas para explorar y generar más participación.
4) Involucre a las personas directamente @mencionándolas por su nombre y preguntándoles específicamente. Si puede, señale cualquier resultado claro y compartido o la comprensión de esos resultados.

   Consejo 8. Al igual que con cualquier hilo, un título claro y simple ayudará a las personas a encontrar respuestas para participar ahora y reflexionar en el futuro.

5) Al programar le recomendamos utilizar una encuesta de tiempo para recopilar respuestas más fácilmente y ver en qué tiempos se responden.

¿CÓMO ANALIZO SUS RESULTADOS?

Para finalizar la aplicación del debate es necesario que se efectúe una evaluación del trabajo realizado. Para ello, se sugiere elaborar un Informe de Participación, en que se presenten los argumentos a favor y en contra del tema debatido. Además, el moderador, el evaluador y el sintetizador (si se decidió la existencia de estas dos últimas entidades) deben entregar por separado su síntesis de las observaciones del debate, destacando las principales debilidades y fortalezas de la actividad, ateniéndose especialmente a la elaboración de argumentos llevada a cabo por los expositores.

De igual forma le recomendamos retomar los resultados del o los instrumento(s) que acompañaron la realización del debate. En este capítulo le ofrecemos un ejemplo de dos instrumentos, una rúbrica y una lista de cotejo. La primera se enfoca más en el contenido, es decir en la calidad de la argumentación en el debate, mientras que la segunda ayuda a valorar el comportamiento y seguimiento de reglas durante el debate (ver Tablas 1 y 2 respectivamente).

¿Cómo lo analizo en línea?

Si la aplicación del debate se llevó a cabo en línea el profesor-facilitador debe tomar en cuenta, además de lo que se incluye en el apartado anterior, lo siguiente:

- Apegarse a los criterios que estableció desde la planeación del debate ya que al igual que en la aplicación presencial es de vital importancia tener claro el propósito del debate y lo que se espera en el desempeño de los alumnos durante el mismo.
- La retroalimentación se puede realizar de manera individual o grupal de manera sincrónica o asincrónica, dependiendo del tipo de plataforma que se emplee.
- El análisis dependerá de la información que provea la plataforma, sin embargo, se puede hacer un Informe de participación con base en los resultados del o los instrumento(s) que acompañaron la realización del debate.

EJEMPLOS

Tabla 1. Rúbrica holística para valorar la calidad de la argumentación en el debate

Asignatura:
Grupo:
Nombre del debate:
Fecha:

Instrucciones: Utilice los criterios descritos a continuación para valorar si el debate cumplió con ellos o no. Señale con una ✔ el recuadro que corresponda a su valoración. Si tiene algún comentario puede anotarlo en el recuadro de observaciones al final del documento.

Escala	Descriptor
Sobresaliente	Se realizó una búsqueda amplia de información asociada a la posición que se defiende. Se discriminó la información relevante tomando en cuenta a la audiencia y al jurado que evaluará a los participantes. Se llevó a cabo un análisis de las tesis fundamentales de las posiciones del debate: principales, secundarias y complementarias. Los argumentos se plantearon con base en afirmaciones, razonamientos y evidencias. En la etapa de "ataques" se defendieron efectivamente las posturas para anular y quitarle valor a los argumentos contrarios a través de la objeción de ideas, términos, exposición de inconsistencias. La exposición de argumentos tuvo rasgos de empatía, el discurso fue creíble, sencillo y retuvo la atención del público. El uso de la voz ayudó a la exposición clara de los argumentos cuidando la dicción, la fluidez, la modulación y una velocidad adecuada. El uso del cuerpo y de los gestos enfatizó fuertemente la argumentación demostrando dominio escénico.
Bueno	Se realizó una búsqueda moderada de información asociada a la posición que se defiende. Se discriminó la información relevante tomando en cuenta solo a la audiencia y no al jurado que evaluará a los participantes. Se llevó a cabo un análisis superficial de las tesis fundamentales de las posiciones del debate: principales, secundarias y complementarias. Los argumentos se plantearon solo con base en afirmaciones y razonamientos, sin evidencias. En la etapa de "ataques" se defendieron débilmente las posturas para anular y quitarle valor a los argumentos contrarios a través de la objeción de ideas y términos sin exposición de inconsistencias. La exposición de argumentos tuvo rasgos de empatía, el discurso fue creíble y sencillo, pero perdió la atención del público. El uso de la voz ayudó a la exposición clara de los argumentos cuidando la dicción, la fluidez, la modulación, pero con una velocidad no adecuada. El uso del cuerpo y de los gestos enfatizó débilmente la argumentación, pero se demostró algo de dominio escénico.

Regular	Se realizó una búsqueda breve de información asociada a la posición que se defiende. Se discriminó la información relevante sin tomar en cuenta a la audiencia ni al jurado que evaluará a los participantes. Se llevó a cabo un análisis pobre de las tesis fundamentales de las posiciones del debate. Los argumentos se plantearon solo con base en afirmaciones, sin razonamientos ni evidencias. En la etapa de "ataques" se defendieron pobremente las posturas para anular y quitarle valor a los argumentos contrarios solo a través de la objeción de ideas sin objeción de términos ni de exposición de inconsistencias. La exposición de argumentos tuvo rasgos de empatía, pero el discurso no fue creíble ni sencillo por lo que perdió la atención del público. El uso de la voz ayudó a la exposición clara de los argumentos cuidando la dicción, la fluidez, pero con mala modulación y sin una velocidad adecuada. El uso del cuerpo y de los gestos enfatizó pobremente la argumentación con un poco de dominio escénico.
Deficiente	No se buscó información asociada a la posición que se defiende. No se discriminó la información relevante. No se llevó a cabo un análisis de las tesis fundamentales de las posiciones del debate. Los argumentos no se plantearon con base en afirmaciones, razonamientos ni evidencias. En la etapa de "ataques" no se defendieron las posturas para anular y quitarle valor a los argumentos contrarios. La exposición de argumentos no tuvo rasgos de empatía, el discurso no fue creíble ni sencillo por lo que se perdió la atención del público. El uso de la voz no ayudó a la exposición clara de los argumentos porque se descuidó la dicción, la fluidez, la modulación y la velocidad adecuada. El uso del cuerpo y de los gestos no enfatizó la argumentación demostrando falta total de dominio escénico.
Observaciones	
Valoración	

Tabla 2. Lista de cotejo para valorar el seguimiento de las reglas y del comportamiento esperado durante el debate formal

Asignatura:
Grupo:
Nombre del debate:
Fecha:

Instrucciones: Utilice los criterios descritos a continuación para valorar si el debate cumplió con ellos o no. Señale con una X el recuadro que corresponda a su valoración. Si tiene algún comentario u observación puede anotarlo en el recuadro de observaciones o bien al final del documento.

Criterio	Sí	No	Observaciones
El moderador empleó su autoridad a lo largo del debate para establecer las condiciones más justas, equitativas y respetuosas posibles para el despliegue de las ideas.			
Se respetó el derecho a réplica con la autorización de relevo del moderador.			
Se respetó el tiempo de cada uno. Las intervenciones fueron lo más breves y concretas posibles y no acapararon el tiempo ni impidieron al otro expresarse.			
En ningún momento dos o más personas hablaron al mismo tiempo.			
Los participantes se apegaron al tema del debate, y no lo abandonarlo a favor de otros que son más afines o convenientes.			
Los participantes hablaron con seguridad y libertad.			
Todos escucharon con atención.			
Todos emplearon un vocabulario adecuado, evitando palabras de registro informal o altisonantes.			
El debate transcurrió de manera organizada, pacífica y respetuosa, exento de argumentos ***ad hominem*** y de ejercicios de violencia física o psicológica.			
Observaciones generales			

RECURSOS EN LÍNEA

Tabla 3. Recursos en línea

Recursos	Descripción
Stack Exchange https://stackexchange.com/	La plataforma permite a los participantes valorar las preguntas y respuestas publicadas y con ello consigue crear foros automoderados. Los participantes pueden conseguir reconocimiento público gracias a la calidad de sus respuestas.
Loomio https://www.loomio.com/	En Loomio los grupos, pueden ser públicos o privados, y sirven para crear discusiones sobre temas específicos. Los miembros autorizados pueden publicar comentarios y crear propuestas, o mejorar propuestas existentes. Finalmente, las propuestas se someten a votación y los miembros pueden estar de acuerdo, en desacuerdo, abstenerse o bloquear, donde el bloqueo se entiende como una forma dura de estar en desacuerdo.
Disqus https://disqus.com/	Los usuarios registrados en la plataforma pueden dejar comentarios en las páginas que utilicen este sistema y sus comentarios pueden ser valorados por otros. Si un usuario deja un comentario en un sitio web y el administrador lo identifica como "spam", ya no podrá dejar más comentarios en otros sitios que utilicen esta plataforma. Este sistema de calidad hace que los usuarios pongan mayor atención en los comentarios que dejan.
Deeyoon https://www.deeyoon.com/	Ofrece la posibilidad de realizar debates en Internet usando el vídeo como protagonista. La identificación es vía Facebook y se puede comenzar a subir vídeos opinando sobre cualquier tema. Se pueden ver los debates ya registrados y votar por la opinión mostrada en cada parte, permitiendo obtener números estadísticos relacionados con la voz popular. De momento solo se encuentran debates en inglés.
Moodle https://moodle.org/	Moodle es una herramienta de gestión de aprendizaje o más concretamente de Learning Content Management de distribución libre. Cuenta con una opción que se llama "foro", sabiéndolo utilizar es muy práctico ya que no implica costos extra y puede contribuir a llevar a cabo un debate entre todos los integrantes de un equipo.

CONCLUSIONES Y REFLEXIONES FINALES

Todo debate tiene el cometido fundamental de contrastar los puntos de vista disponibles en una materia, a través de dos o más expositores, de manera seria, argumentada y frontal, de modo tal que quienes asistan al debate puedan recibir información pertinente y puedan hacerse una opinión propia. Esto significa que el cometido del debate no es realmente ganar, pues no se trata de un concurso, si bien cada participante se esforzará por convencer lo más posible a los demás de su punto de vista, como es normal (Raffino, M., 2021).

El debate permite que los estudiantes activen procesos cognitivos asociados a la organización de la información, tanto interna como externa. Se utiliza durante el momento de desarrollo de la clase, con el fin de enfrentar una situación desde distintas perspectivas, fomentando la argumentación y discusión. Cuando se quiere potenciar en los estudiantes el desarrollo de competencias que requieren: pensamiento crítico y lógico, trabajo en equipo, uso de recursos del lenguaje y de la comunicación no verbal. Además, prepara a los estudiantes para enfrentar situaciones de presión y para la defensa argumentada de ideas y planteamientos, ya que deben seleccionar información desde diversas fuentes fiables y aplicarlas en la estructuración de una argumentación.

Estimula el proceso de aprendizaje y ayuda a que exista una buena dinámica grupal.

En el salón de clases, los debates suelen realizarse con frecuencia, aunque con características menos formales. Por ejemplo, el público, los participantes y el moderador no están siempre definidos; sin embargo, este género oral es una forma de alcanzar la participación activa de todos los integrantes del grupo, la cual no se obtiene con la exposición oral, por ejemplo. Además, el debate permite ver y contrastar posturas opuestas, ampliando así la perspectiva de los estudiantes.

El debate es una herramienta pedagógica de utilidad cuando se busca desarrollar habilidades críticas, así como ampliar y facilitar la comprensión de un tema determinado. Este instrumento fomenta el pensamiento crítico, así como el respeto y la tolerancia ante opiniones diferentes. Además, es una herramienta muy útil para transmitir conocimiento, de manera rápida y activa.

Ejercicios interactivos

Recomendamos realizar el siguiente ejercicio interactivo que consiste en responder un quiz sobre los elementos más importantes del debate abordados en este capítulo. El enlace es https://docs.google.com/forms/d/e/1FAIpQLSdm-RDsCf9d3M5fRgLgFNE8kDgQrDgMp_-ZmVNViGjurXzLww/viewform?usp=sf_link

REFERENCIAS

Argudín, M. (s.f.). Evaluación / instrumentos centrados en el alumno. Disponible en: http://hadoc.azc.uam.mx/evaluacion/debate.htm. Consultado el 01 en abril de 2021

Raffino, M. (2021) *Debate* en Concepto de Argentina. Consultado el 01 de abril de 2021 en https://concepto.de/debate/#ixzz6qpSloe8l

Centro de escritura y pensamiento crítico (1 de abril de 2015). ¿Sabes cuáles son las características de un debate? *Blog de la Universidad de Las Américas Puebla*. http://blog.udlap.mx/blog/2015/04/caracteristicasdeldebate/#:~:text=La%20estructura%20del%20debate%20est%C3%A1,y%20respuestas%2C%20y%20la%20conclusi%C3%B3n.

CriticaMente. (2018. 5 de agosto) *Formatos de debate.* [Video]. YouTube. https://www.youtube.com/watch?v=tF50jbdYAXI

Fundación educativa activa-t. (2014, 23 de enero) *Debate: Estructura y organización de las ideas.* [Video]. YouTube. https://www.youtube.com/watch?v=USqZZ2QbT5Y

Moneo, A. (2015). *6 herramientas para facilitar los debates en internet.* Disponible en: https://blogs.iadb.org/conocimiento-abierto/es/6-herramientas-para-facilitar-los-debates-en-internet/ Consultado el 07 de octubre de 2021.

Capítulo 34

EXAMEN CLÍNICO OBJETIVO ESTRUCTURADO (ECOE)

Guadalupe Soto Estrada, Adrián Israel Martínez Franco, Adrián Martínez González

"En la medida en que un sujeto aprende, simultáneamente evalúa, porque discrimina, valora, critica, opina, razona, fundamenta, decide, enjuicia, opta… entre lo que considera que tiene un valor en sí y aquello que carece de él. Esta actividad evaluadora, que se aprende, es parte del proceso educativo, que como tal es continuamente formativo."

Pedro Ahumada Acevedo

INTRODUCCIÓN

La enseñanza de las habilidades clínicas que conforman la competencia clínica es un elemento indispensable para el aprendizaje del estudiante de Medicina, Enfermería, Odontología, Psicología, Fisioterapia, entre otras, ciencias de la salud y la evaluación de esta resulta de mucha utilidad para valorar el desarrollo y desempeño de los estudiantes de las ciencias de la salud, tanto en la modalidad presencial como a distancia.

En las tres últimas décadas, el Examen Clínico Objetivo Estructurado (ECOE) se ha convertido en una herramienta indispensable para evaluar el aprendizaje de los alumnos en la práctica clínica, la cual se evaluaba con el examen oral y el examen ante paciente real. El ECOE tiene amplias evidencias de validez en numerosos estudios, incluyendo uno realizado en la Facultad de Medicina de la UNAM (Trejo-Mejía, Sánchez-Mendiola, Méndez-Ramírez, y Martínez-González, 2016). En 1975, Harden publicó la versión original del ECOE y desde entonces se han ido modificando algunas características de este examen, con la finalidad de mejorar la calidad con la que se evalúa a los alumnos.

Habitualmente, el ECOE se emplea formativamente en estudiantes de las ciencias de la salud de pregrado y posgrado e incluso en los exámenes de certificación de especialistas y se considera cada vez más, un elemento imprescindible dentro de la evaluación sumativa para que el alumno pueda obtener el título, el grado o la certificación, según sea el caso. Es conveniente mencionar que este examen, con las adecuaciones pertinentes, también se está utilizando en otras disciplinas como en Derecho, en Sociología, en Trabajo Social, entre otras (Bogo et al., 2012; Espinosa-Vázquez, Sánchez-Mendiola, Leenen, y Martínez-González, 2020; Manuel, 2010).

Este capítulo está organizado en ocho secciones. En la primera sección bajo el título ¿Qué es?, se describen las características de esta estrategia describiendo los elementos más

importantes que lo conforman, en la segunda sección se explican cuáles son las pautas para su diseño en modalidad presencial o a distancia, en la tercera sección se desarrolla su aplicación, en la cuarta sección se describe su aplicación en la evaluación a distancia, en la quinta sección se dan ciertas recomendaciones para analizar sus resultados, en la sexta sección se ofreces varios ejemplos con los formatos que se utilizan en una estación dinámica, en la séptima sección se da una lista de recursos en línea para el ECOE y finalmente, en la octava sección se dan conclusiones y recomendaciones.

¿QUÉ ES?

- El Examen Clínico Objetivo Estructurado (ECOE) representa un encuentro entre un paciente que simula una enfermedad predeterminada (caso clínico) y el estudiante de pregrado o posgrado o médico que será evaluado en un escenario controlado (estación).
- El ECOE es una estrategia de evaluación del aprendizaje que contribuye a la estimación del desempeño integral de los estudiantes, es decir, de la competencia clínica entendida como el conjunto de conocimientos, habilidades, actitudes y valores que, interrelacionados entre sí y con base en el razonamiento, la evidencia científica y el juicio crítico, permiten un desempeño profesional eficiente (Martínez-González, Sánchez-Mendiola, Méndez-Ramírez, y Trejo-Mejía, 2016). Se considera el estándar de oro para evaluar la competencia clínica.
- Algunos de los atributos de la competencia clínica que se evalúan con esta estrategia son: la entrevista o interrogatorio, la exploración física, la interpretación de los estudios de laboratorio y gabinete, el diagnóstico y plan de manejo, así como la comunicación en la relación médico-paciente (Durante, Lozano, Martinez, Morales, y Sanchez, 2012; Neufeld y Norman, 1985).
- La finalidad de cada estación es evaluar ciertos conocimientos, habilidades, actitudes y valores adecuados al nivel de estudios del examinado. Para esto se puede emplear un paciente real o simulado, ambos estandarizados porque representan el mismo papel en repetidas ocasiones, con un padecimiento predeterminado en un caso clínico y en una estación dinámica o estática.
- En cada estación se plantea uno de los múltiples escenarios posibles que el futuro profesional enfrentará en su práctica diaria. En las estaciones dinámicas se pueden evaluar diversas actividades como son: el interrogatorio, la exploración física, el diagnóstico o la terapéutica de un paciente, es decir, hay una relación interpersonal paciente-examinado.
- También se pueden instalar estaciones denominadas estáticas en las que no hay paciente, sino que están dirigidas a que el estudiante demuestre otro tipo de habilidades como son: la interpretación de pruebas de laboratorio, interpretación radiográfica o del electrocardiograma, la búsqueda en bases de datos o un análisis de trabajos de investigación.
- En cada estación se requiere un evaluador, quien, a partir de una herramienta de evaluación, generalmente una lista de cotejo o rúbrica, determina el resultado de la evaluación, el cual debe considerarse como una oportunidad para el aprendizaje al realimentar de forma efectiva al estudiante.

Tipos de estaciones:

- **De procedimientos.** En ellas, el estudiante interactúa con un paciente real, simulado, estandarizado o con un maniquí para realizar una tarea específica como son: el interrogatorio, la exploración física, el diagnóstico y/o el manejo integral.
- **De interpretación de estudios de laboratorio o gabinete.** El estudiante, de acuerdo con un caso clínico, debe anotar los hallazgos y establecer un diagnóstico.
- **De análisis de estudio de investigación**. En ellas, el estudiante, al recibir el resumen de un trabajo de investigación, revisa los aspectos metodológicos para valorar sus resultados o interpretarlos para la toma de decisiones.

¿CÓMO LO DISEÑO?

El ECOE es una prueba en la que existe cierto número de estaciones, las cuales representan lo equivalente a un consultorio o al área de urgencias de un hospital, por lo regular son entre 9 y 18 estaciones conectadas a través de un circuito, por medio del cual los estudiantes evaluados pasan uno por uno.

Por lo tanto, un elemento fundamental para llevar a cabo un ECOE, es el desarrollo de casos para las estaciones por las que tendrá que pasar el examinado para ser evaluado. Los casos se desarrollan con base en los conocimientos, habilidades, actitudes y procedimientos que se desean evaluar, los cuales se obtienen considerando el perfil de egreso que se desea lograr y los programas de estudio. Se sugiere utilizar la información de expedientes de casos reales.

Antes de iniciar con la elaboración de las estaciones, es necesario contar con una tabla de especificaciones y una matriz de contenidos, esto con la finalidad de que se señalen claramente las áreas, los conocimientos, las habilidades y actitudes que se evaluarán de acuerdo con el perfil del egresado. Dichos documentos se deben elaborar con la participación de un cuerpo colegiado o expertos del área que será evaluada. Esto permitirá un balance entre los contenidos del ECOE y el peso que tienen en su programa de estudio vigente.

Figura 1. Circuito de estaciones en un ECOE

Estación de descanso
10
1
2
3
4
5
6
7
8
9
ECOE
• 15m en cada estación
• 10 estaciones en 2h 30m

Planeación

Es importante la planeación anticipada del proceso logístico y operativo del examen, así como determinar los atributos de la competencia clínica esperada en el evaluado según el programa académico por medio de una tabla de especificaciones y una matriz de contenidos, esto ayudará a seleccionar a los evaluadores y pacientes quienes deben ser capacitados y familiarizados en el ECOE.

Objetivo de la evaluación

Describir qué atributos de la competencia clínica se desean evaluar. Es útil definir qué se va a evaluar: ¿la competencia clínica?, cómo se va a evaluar: ¿a través de un ECOE?, ¿cuándo?, ¿a quiénes? Para qué: ¿será una evaluación diagnóstica, formativa, sumativa o todas ellas?

Lugar de la evaluación

Definir el lugar donde se instalarán las estaciones del ECOE para llevar a cabo la evaluación del examinado. Es importante hacer un registro tanto del material como de los recursos que se requieren para llevar a cabo el examen de acuerdo al número de examinados. Es conveniente definir el número de evaluadores y pacientes que van a intervenir en el examen, así como el de examinados y el tiempo a utilizar en cada estación como en el ECOE global.

Tipo de escenario

Definir qué tipo de estación se le planteará al examinado, ya sean procedimientos (dinámica) o de interpretación de estudios de laboratorio y gabinete (estática).

Tipo de estación

De acuerdo con el tipo de estación establecido se deberá planear la logística; el lugar que se necesita y si la estación podrá evaluar los objetivos planteados.

Elaborar y preparar las estaciones

Las estaciones dinámicas requieren de tres documentos que deben ser elaborados por los profesores expertos en el área:

- *Formato del estudiante:* Incluye la presentación del caso y las instrucciones que especifican lo que se espera realice el estudiante (Ver en la sección de ejemplos: Ejemplo 1).
- *Libreto del paciente estandarizado:* Incluye toda la información clave que necesita memorizar respecto a sus características generales y a su padecimiento actual, la cual debe utilizar para responder a las preguntas que le haga el estudiante (Ver en la sección de ejemplos: Tabla 2).
- *Formato del evaluador:* Contiene el instrumento con el que se evaluará el desempeño del alumno y que puede ser una lista de cotejo o una rúbrica (Ver en la sección de ejemplos: Tabla 3).

En las estaciones estáticas se utilizan solamente el formato del estudiante y el formato del evaluador.

Otros documentos que se pueden necesitar pueden ser los resultados tanto de laboratorio o gabinete como de citología o datos de exploración física, mismos que deben incluirse para que el examinado pueda lograr el objetivo de la estación.

Selección del material e instrumento de evaluación

Una vez definidos los procedimientos que se desea evaluar y el espacio físico apropiado, lo más parecido a un escenario real, si se trata de una estación dinámica se debe considerar que exista un paciente capacitado o modelo y una mesa de exploración, asimismo, se debe contar con los instrumentos necesarios para realizar el procedimiento solicitado, como estuche de diagnóstico, vendas, jeringas, entre otros.

Cuando se trata de una estación estática se debe verificar que se cuente con el material necesario, pruebas de laboratorio y gabinete, microscopio, computadora, lápiz y papel, entre otros.

Seleccionar el instrumento de evaluación (puede ser una rúbrica o una lista de cotejo) que nos permitan conocer el nivel de desempeño de cada sustentante. Debe contener los indicadores que se quieran evaluar y que correspondan con los objetivos planteados. Se debe seleccionar un instrumento con fuentes de evidencia de validez como la rúbrica (Martínez-González et al., 2020) y de ser posible una tableta para recolectar la información del desempeño del examinado o el formato en papel.

Es de suma importancia cuidar la congruencia entre el caso clínico, las instrucciones para el estudiante, el instrumento de evaluación que se empleará y el libreto del paciente.

Se sugiere que las estaciones estén cercanas y se encuentren ubicadas en una secuencia lógica y fácil de seguir para los estudiantes.

¿CÓMO LO APLICO?

- Es necesario realizar una prueba piloto con la finalidad de identificar errores y corregirlos antes de que sea aplicado de manera formal a los examinados, por ejemplo, en el libreto del paciente, preguntas mal redactadas, mal planteadas o confusas.
- Aplicar el ECOE sumativo, formativo o diagnóstico a los examinados. Se deberá evaluar a todos los examinados con el mismo instrumento y registrar los resultados en una base de datos para analizarla posteriormente.
- Los resultados de la prueba se deben analizar por el cuerpo de profesores o bien por los integrantes del comité de evaluación, para que posteriormente se den a conocer los resultados a los examinados y a las autoridades correspondientes. Más adelante, se profundizará sobre el análisis de los resultados.

Sugerencias

Es de suma importancia que se dé retroalimentación a los examinados, ya que la evaluación formativa representa la gran fortaleza para el aprendizaje de las habilidades clínicas, que comprenden la competencia clínica, para que ellos puedan identificar sus áreas de oportunidad y tengan la posibilidad de corregir los errores y así poder mejorar su práctica profesional.

El proceso que realiza el examinado en la estación puede evaluarse mediante una rúbrica o una lista de cotejo.

EL ECOE EN LA EVALUACIÓN A DISTANCIA

¿Cómo lo diseño a distancia?

Para el diseño del ECOE a distancia la planeación es similar al ECOE presencial, los atributos centrales del ECOE que se pueden evaluar son: las habilidades de comunicación oral y

escrita, habilidades prácticas, de diagnóstico, tratamiento y profesionalismo, sin embargo, hay evidencia de que las habilidades prácticas tienen gran dificultad para ser evaluadas en la modalidad a distancia; aún con esa limitación, hay experiencias que han comenzado a aparecer a nivel internacional e incluso en México con el ECOE a distancia también llamado Web ECOE (Hannon, Lappe, Griffin, Roussel, y Colbert-Getz, 2020; Major, Sawan, Vognsen, y Jabre, 2020; Rivero-López et al., 2021) o ECOE virtual, por lo que a continuación ofrecemos algunas sugerencias prácticas para quienes deseen utilizar esta modalidad.

- **Sugerencia 1. Planeación del ECOE virtual.**
  El ECOE virtual es un concepto nuevo para la mayoría de las instituciones por lo que necesita dedicar tiempo para su planeación. Esto incluye conformar y capacitar, en su caso, un cuerpo colegiado de académicos quienes deben considerar el propósito, el número de estudiantes, las habilidades a evaluar, las estaciones a incluir, el circuito, los recursos disponibles y el tiempo. La duración del examen puede ser de dos a tres horas distribuidas de la siguiente forma cuando se empleen 30 minutos por estación en un ECOE formativo: 20 minutos para la entrevista, 2 minutos de retroalimentación del paciente, 3 minutos de retroalimentación del evaluador y 5 minutos para cambio de estación y descanso.

  Otro punto para considerar, si se trata de un ECOE a distancia, es que, tanto el personal de apoyo como los evaluadores y los sustentantes cuenten con el equipo y la conexión a internet adecuados para llevar a cabo el examen.
- **Sugerencia 2. Elija la plataforma tecnológica más adecuada para su ECOE virtual.**
  Hay varias plataformas en línea disponibles en las que se puede ejecutar su ECOE virtual, incluidos Microsoft Teams, Zoom, Google Hangouts y Adobe Connect. Cada plataforma tiene pros y contras, y se actualizan repetidamente con nuevas funciones, por lo que recomendamos elegir una plataforma con la que los estudiantes y profesores estén más familiarizados. El soporte tecnológico (TI) a lo largo de este proceso es fundamental para su éxito. Algunos artículos describen la importancia de colaborar con el departamento de Tecnologías para el hardware, software y sistemas basados en la nube necesarios para el ECOE virtual. Lara, Foster, Hawks, y Montgomery, 2020; Prettyman, Knight, y Allison, 2018; Rivero-López et al., 2021 utilizaron Zoom de manera efectiva. Estos autores encontraron útil la función de sesión de grupos pequeños, ya que les permitía a los estudiantes rotar de la "sala de espera" a una "sala de examen" virtual separada donde el paciente (actor) y el examinador ya estaban ubicados. Es conveniente considerar que una persona del departamento de tecnologías esté presente durante el examen para solucionar cualquier incidente que pueda surgir. Con fines formativos, merece la pena considerar un software que pueda grabar estaciones fácilmente, brindando a los estudiantes la oportunidad de ver y aprender de su desempeño (Prettyman et al., 2018). Esto también puede ser útil para la formación de examinadores y el desarrollo de más estaciones del ECOE virtual.

- **Sugerencia 3. Seleccione y modifique sus estaciones para un ambiente virtual.**
  Con base en las competencias a evaluar y la tabla de especificaciones, en caso de contar con estaciones elaboradas para la modalidad presencial puede adaptarlas para aplicarlas en un ambiente virtual. Por supuesto, también se pueden generar nuevas estaciones por el cuerpo colegiado a fin de ir acumulando fuentes de evidencia de validez de éstas.
- **Sugerencia 4. Considere los recursos materiales para cada estación.**
  Las habilidades prácticas pueden ser evaluadas con narraciones donde los evaluados mencionan cómo realizarían la exploración y qué podrían encontrar en el paciente. En caso de que sea factible demostrar habilidades prácticas, por ejemplo, suturar en telas o en otro material deben encontrar la mejor área y el mejor ángulo de cámara para que el examinador visualice completamente su desempeño. No se pueden solicitar nuevos materiales de último minuto, ya que esto alertaría a los estudiantes sobre lo que se evaluará. Para el día del ECOE, los estudiantes deberán reunir todo el material y equipo con anticipación, por lo que deben contar de forma anticipada con una lista de lo que podría ser necesario. Se sugiere incluir recursos diferentes a los que utilizará para sus estaciones. Esto evita advertir a los estudiantes sobre qué habilidades se evaluarán.
- **Sugerencia 5 Capacitación para un ECOE virtual.**
  Todos los participantes deberán familiarizarse con el ECOE virtual y con la plataforma tecnológica utilizada con tiempo de anticipación. Tanto los evaluadores que utilizarán el instrumento de evaluación, como los pacientes que serán estandarizados para que sean consistentes en sus respuestas, además, los estudiantes deben ser capacitados adecuadamente para garantizar que su desempeño no se vea afectado por la modalidad de evaluación y en cambio, sí se evalúe realmente su competencia clínica. Es conveniente que las personas que participarán como evaluadores hayan tenido alguna experiencia en ECOE presenciales.
- **Sugerencia 6. Mantenga la evaluación centrada en el estudiante.**
  Si la finalidad del ECOE es formativa, al final de cada estación el evaluador y el paciente pueden dar retroalimentación al estudiante. La retroalimentación efectiva tiene un gran potencial para el aprendizaje del estudiante cuando es utilizada de manera oportuna e individualizada.
- **Sugerencia 7. Optimice lo que se mostrará en pantalla.**
  Para cada estación, el evaluador debe ver claramente a su candidato y de preferencia utilizar una tableta para recolectar la información sobre el desempeño del estudiante en la rúbrica.

  Si se necesitan documentos adicionales durante una estación, el evaluador puede compartir su pantalla para mostrar este documento a solicitud del examinado (por ejemplo, biometría hemática, química sanguínea, entre otros).

  Los examinadores deben tener su video encendido inicialmente, para saludar y verificar que los examinados puedan iniciar las estaciones sin contratiempos, y posteriormente apagar su video, el cual debe de encender cuando el examinado solicite alguna prueba de laboratorio o gabinete y mostrarla en pantalla, volver a apagarlo y al final de la estación encenderlo nuevamente para realimentar al estudiante.

- **Sugerencia 8. Realice una prueba piloto.**
  Se recomienda que antes de la aplicación de ECOE en línea se realicen ensayos previos al examen para conocer la dinámica y afinar detalles o problemas que surjan durante el proceso. La prueba piloto es útil para garantizar la factibilidad de las estaciones, pueden colaborar estudiantes voluntarios que no intervendrán en el ECOE. Los evaluadores y pacientes estarán más seguros de su rol y podrán comprobar que se tienen todos los recursos y se puede identificar cualquier problema tecnológico. Por ejemplo, algunas marcas de computadora no muestran el formato correcto de la documentación cuando se comparte con una laptop. Durante la prueba piloto, reserve el tiempo suficiente para darle a cada examinador y paciente la oportunidad de ingresar y salir de la estación, probar su micrófono y practicar la activación y desactivación de audio y video.
- **Sugerencia 9. Utilice un canal de comunicación privado para asegurarse que los examinados no puedan acceder al material antes del ECOE.**
  Es útil tener un canal externo dedicado para la comunicación entre organizadores, asegurando que el ECOE real no se vea perturbado. Puede utilizar un grupo de Telegram o WhatsApp para una comunicación rápida entre diferentes miembros del equipo involucrado en el ECOE. Los estudiantes también pueden utilizar WhatsApp para poder contactar al evaluador en cualquier momento en caso de problemas de conectividad.
- **Sugerencia 10. Obtenga retroalimentación de todos los participantes.**
  Al finalizar el examen obtenga comentarios de los examinados, pacientes y evaluadores y aprenda de ellos para mejorar el siguiente ECOE virtual.

La aplicación de un ECOE virtual es factible siguiendo las sugerencias anteriores y brinda la oportunidad de probar una amplia gama de habilidades clínicas en un entorno a distancia. Requiere planificación, capacitación y prueba piloto para garantizar que todos los involucrados conozcan sus funciones y estén familiarizados con el formato y el software. Este cambio en la práctica establecida brinda oportunidades para la evaluación de habilidades clínicas durante esta situación de pandemia y la evaluación futura de los estudiantes en ubicaciones a distancia.

Video sugerido:

En el siguiente video se presenta la conferencia “Experiencia de la aplicación del ECOE en línea en el Examen Profesional” en el marco del Primer Congreso Internacional de Tecnología para la Educación Médica: https://youtu.be/LRREEruo9jg

Ejercicio interactivo:

Lo invitamos a que con el apoyo del siguiente video de la Universidad de Murcia: https://youtu.be/5wtUvonW3zE elabore una propuesta de caso de su área de experiencia como docente para ser aplicado en un ECOE ya sea presencial o a distancia.

¿CÓMO ANALIZO LOS RESULTADOS?

Para analizar los resultados obtenidos a partir de un ECOE se debe tomar en cuenta lo siguiente:

- Es necesario recabar las opiniones sobre el instrumento, el proceso y el contenido del examen, tanto del profesor como del estudiante, esto con la finalidad de corregir posibles fallas y errores al evaluar el desempeño de los estudiantes en las estaciones.
- Por lo regular, a partir de la base de datos se pueden obtener medidas de resumen, por ejemplo, se puede registrar el puntaje obtenido por estudiante y obtener medidas de tendencia central y de dispersión.
- El análisis debe considerar la medición de la confiabilidad para estimar la validez interna del examen aplicado, es decir, se necesita estimar la reproducibilidad del examen a lo largo del tiempo o en múltiples ocasiones (consistencia). Para esto se puede utilizar inicialmente la prueba de análisis de varianza para ajustar el modelo. Esto puede ser mediante un modelo de efectos aleatorios. De acuerdo con (J. A. Trejo-Mejía et al., 2016), este tipo de modelo permite obtener los estimadores de los componentes de la varianza para cada variable que se haya considerado. Además, es importante tomar en cuenta que el error de medición siempre ocurre y el nivel en que el puntaje de cada individuo evaluado puede variar se denomina coeficiente de confiabilidad y éste se puede calcular mediante varias herramientas, tales como el Alfa de Cronbach o la teoría de la generalizabilidad.
- A partir de los resultados obtenidos, tanto de forma cuantitativa como cualitativa, es útil elaborar un reporte donde el estudiante pueda revisar su desempeño en el examen. Es conveniente que el profesor tenga una sesión con sus estudiantes en la que se comenten los resultados y se den las conclusiones acerca del ECOE realizado, así como una retroalimentación general que incluya las fortalezas de los examinados, así como las principales áreas de oportunidad.

EJEMPLOS

A continuación, se presenta un ejemplo del material que se elabora para una estación dinámica que será evaluada dentro de un ECOE de varias estaciones:

Ejemplo 1
Instrucciones para el estudiante que será evaluado

Mujer de 27 años de edad, que acude al servicio de urgencias por presentar dolor abdominal.

Signos vitales:

- Temperatura: 37.5ºC
- Presión arterial: 110/70 mm Hg

- Pulso: 82/min
- Frecuencia respiratoria: 18/min

Instrucciones para el estudiante

1. Obtener la información pertinente sobre el problema del paciente y los antecedentes relevantes.
2. Realizar un examen físico específico al padecimiento (no realizar un examen de mama, pélvico/genital, reflejo corneal ni rectal).
3. Discutir sus impresiones y planes iniciales con el paciente.
4. Al terminar, complete la nota del paciente en el formato entregado.

Ejemplo 2

Tabla 1. Libreto del paciente, preguntas y respuestas

Contestar a los alumnos solo lo que le pregunten	
¿Cuál es su nombre?	Diga su nombre
¿Cuál es su edad?	27 años
¿Por qué acude a consulta?	Tengo dolor en la parte baja de mi estómago
¿Desde cuándo tiene el dolor?	Hace aproximadamente una hora
¿Cómo comenzó?	Repentinamente
¿Qué estaba haciendo?	Estaba sentada viendo la TV
¿Ya había tenido este dolor antes?	No, es la primera vez
¿Cómo es ese dolor?	No entiendo doctor
¿Es como ardor, como opresión, como cólico?	Es como cólico menstrual, pero más fuerte
¿Dónde le duele?	Del lado derecho abajo del estómago
¿Se le corre para algún lado?	No está solo ahí
¿Cómo califica la intensidad de su dolor?	De 10 me duele 6
¿Con que le disminuye el dolor?	Se calma cuando junto mis rodillas con mi estómago
¿Con que aumenta?	Al extender la piernas, al estar parada o con el movimiento
¿Ha tomado algún medicamento?	No ninguno
¿Tiene náusea o vómito?	Si, náusea leve, sin llegar al vómito
¿Se sabe enferma de algo?	Sí, tengo un quiste en ovario derecho
¿Toma medicina para ello?	Sí, pastillas hormonales
¿Se ha realizado exámenes?	Una prueba positiva de VPH hace 5 años y un Papanicolaou hace seis meses que salió normal

¿Cuántos embarazos ha tenido?	He tratado de embarazarme desde hace tres años sin conseguirlo
¿Tiene algún tratamiento?	Sí, tengo tratamiento de infertilidad con pastillas y desde hace un mes medicina inyectada, algunas veces olvido el tratamiento
¿Cada cuándo menstrua?	Cada 28 días y dura cuatro días normalmente con sangrado moderado con cólicos por lo que tomo ibuprofeno
¿Cuándo fue su última menstruación?	Hace cinco semanas
¿El sangrado fue como de costumbre?	No, fue más ligero y solo 1-2 días
¿Alguna molestia en sus relaciones sexuales?	No
¿Última visita al médico?	Hace dos meses con el especialista de fertilidad para tratamiento. pero este mes no he podido ir por estar muy ocupada
¿Se ha vacunado contra el VPH?	No
¿Familiares con alguna enfermedad?	Una hermana de 30 años de edad con endometriosis
¿A qué se dedica?	Soy aeromoza, llevé un curso específico
¿Cuál es estado civil?	Casada desde hace 3 años
¿Usted fuma o ingiere bebidas alcohólicas?	Fumo y tomo ocasional solo en fiestas.
¿Toma café?	Tres tazas diarias
¿Cómo es su dieta?	La mayor parte como en mi trabajo, dieta normal
¿Realiza ejercicio?	2-3 veces a la semana acudo a clase de spinning
¿Tiene algún pasatiempo?	Viajar, ir de compras o andar en bicicleta de montaña
¿Se ha sentido estresada?	Sí, porque no he podido embarazarme

Fuente: Elaboración propia.

Ejemplo 3
Tabla 2. Guía para el evaluador

La siguiente rúbrica ha sido utilizada en el Examen Clínico Objetivo Estructurado que se aplica como examen profesional a los egresados de la Facultad de Medicina de la UNAM.

Marque la opción elegida con una "X". Ejem.		C. Bueno X	
1. Habilidades en el interrogatorio			
A. Insuficiente Incompleto. Desorganizado. Sin relación con el padecimiento actual del paciente.	B. Suficiente	C. Bueno	**D. Excelente** Muy Completo. Muy organizado. Orientado al padecimiento actual del paciente.

2. Habilidades en la exploración física			
A. Insuficiente Exploración incompleta, desorganizada, sin relación con el padecimiento actual del paciente.	B. Suficiente	C. Bueno	**D. Excelente** Exploración completa, sistemática y orientada al padecimiento actual del paciente.
3. Estudios de laboratorio y gabinete			
A. Insuficiente Solicita e interpreta inadecuadamente los exámenes complementarios de acuerdo al padecimiento actual del paciente.	B. Suficiente	C. Bueno	**D. Excelente** Solicita e interpreta adecuadamente los exámenes complementarios de acuerdo al padecimiento actual del paciente.
4. Habilidades diagnósticas			
A. Insuficiente No identifica los elementos para establecer diagnósticos diferenciales, con base en la evidencia.	B. Suficiente	C. Bueno	**D. Excelente** Identifica los elementos para establecer diagnósticos diferenciales con base en la evidencia.
5. Plan preventivo-terapéutico (médico y/o quirúrgico)			
A. Insuficiente No sugiere el plan preventivo y/o tratamiento adecuado de acuerdo al padecimiento actual.	B. Suficiente	C. Bueno	**D. Excelente** Sugiere el mejor plan preventivo y/o tratamiento de acuerdo al padecimiento actual.
6. Habilidades de comunicación			
A. Insuficiente Problemas graves en la comunicación con el paciente o sus familiares.	B. Suficiente	C. Bueno	**D. Excelente** Se comunica de forma adecuada, respetuosa y efectiva con el paciente o sus familiares.
7. Valoración global de los conocimientos y habilidades			
A. Insuficiente Falta de conocimiento y desarrollo de habilidades de interrogatorio, exploración e interpretación para integrar el diagnóstico y manejo.	B. Suficiente	C. Bueno	**D. Excelente** Sobresaliente conocimiento y desarrollo de habilidades de interrogatorio, exploración e interpretación para integrar el diagnóstico y manejo.
8. Nota médica			
Insuficiente Incompleta. Desordenada. No integra la información para fundamentar los diagnósticos diferenciales.	B. Suficiente	C. Bueno	**D. Excelente** Muy completa y ordenada. Integra adecuadamente la información para fundamentar los diagnósticos diferenciales.
9. Evaluación por el paciente			
A. Insuficiente No saludó, no se presentó ni mostró respeto durante la entrevista o exploración. No utilizó un lenguaje claro ni comprensible.	B. Suficiente	C. Bueno	**D. Excelente** Saludó, se presentó y mostró respeto durante la entrevista y exploración. Utilizó un lenguaje claro y comprensible.

Fuente: Martínez González A, Sánchez Mendiola M, Olivares Olivares SL, et al. (2020). Colaboración de tres escuelas de medicina de México en un examen clínico objetivo estructurado (ECOE). Inv Ed Med; 9: 58-69.

RECURSOS EN LÍNEA

Recurso	Descripción
El ECOE en la UNAM https://youtu.be/IhtyQE3-J_4	En este video se explican los pasos para la aplicación de un ECOE, Facultad de Medicina, UNAM.
Las Fases del ECOE https://www.coursera.org/lecture/evaluacion-cualitativa/ecoe-KrPQj	En este video se explican las características del ECOE, así como las diferentes fases en las que se puede planear y aplicar.
ECOE en la Universidad de Alcalá https://youtu.be/YrZgY8e-SJA	Prueba ECOE en la Facultad de Medicina de la Universidad de Alcalá de Henares, España.
ECOE en posgrado https://www.youtube.com/watch?v=z4mQMuMtBNw&t=34s	ECOE en la especialidad de Otorrinolaringología con fines de Certificación.
Estación dolor abdominal https://youtu.be/HQEZVoBIjZo	Ejemplo de paciente con dolor abdominal en una estación ECOE.
Estación de entrevista https://youtu.be/j4qCALj_cAE	Ejemplo de una estación para evaluar la entrevista en el ECOE.

CONCLUSIONES Y RECOMENDACIONES

- El ECOE se ha convertido en una herramienta de aprendizaje y evaluación muy útil para explorar las habilidades, las destrezas, las actitudes y los valores que han adquirido los estudiantes.
- Para cumplir con los objetivos de evaluación planteados al aplicar un ECOE, es necesario contar con una adecuada planeación y diseño, así como los recursos necesarios para su realización.
- Esta herramienta de evaluación brinda al estudiante la oportunidad de reconocer sus fortalezas y sus áreas de oportunidad a fin de mejorar su práctica profesional.
- Se sugiere aplicar este examen en la educación media superior (bachillerato), en las licenciaturas, especializaciones, maestrías y doctorados en las que se requiera valorar aspectos prácticos que se tendrán que aplicar en el futuro.
- Para revisar un ejemplo particular de cómo realizar un ECOE al inicio y término del internado médico de pregrado (ciclo escolar muy parecido en las escuelas y facultades de medicina) se recomienda consultar (Martínez-González y Trejo-Mejía, 2018; J. Trejo-Mejía, Blee-Sánchez, y Peña-Balderas, 2014).

REFERENCIAS

Bogo, M., Regehr, C., Katz, E., Logie, C., Tufford, L., Litvack, A. (2012). Evaluating an Objective Structured Clinical Examination (OSCE) Adapted for Social Work. *Research on Social Work Practice.*, *22*(4), 428–436. https://doi.org/10.1177/1049731512437557

Durante, I., Lozano, J., Martinez, A., Morales, S., Sanchez, M. (2012). *Evaluación de competencias en ciencias de la salud* (Editorial Médica Panamericana, ed.). UNAM.

Espinosa-Vázquez, O., Sánchez-Mendiola, M., Leenen, I., Martínez-González, A. (2020). Evaluación del desarrollo de la competencia clínica en odontopediatría con el examen clínico objetivo estructurado. *Investigación en Educación Médica*, *9*(34), 53–62. https://doi.org/10.22201/facmed.20075057e.2020.34.19198

Hannan TA, Umar SY, Rob Z & Choudhury RR. (2021). Designing and running an online Objective Structured Clinical Examination (OSCE) on Zoom: A peer-led example, *Medical Teacher*; *43*:651-655, https://doi.org/10.1080/0142159X.2021.1887836

Hannon, P., Lappe, K., Griffin, C., Roussel, D., & Colbert-Getz, J. (2020). An objective structured clinical examination: From examination room to Zoom breakout room. *Medical Education, 54*(9), 861–861. https://doi.org/10.1111/MEDU.14241

Kakadia R, Chen E, Ohyama H. (2020) Implementing an online OSCE during the COVID-19 pandemic. *J Dent Educ*, 1–3. https://doi.org/10.1002/jdd.12323

Lara, S., Foster, C. W., Hawks, M., Montgomery, M. (2020). Remote Assessment of Clinical Skills During COVID-19: A Virtual, High-Stakes, Summative Pediatric Objective Structured Clinical Examination. *Academic Pediatrics*, *20*(6), 760. https://doi.org/10.1016/J.ACAP.2020.05.029

Major, S., Sawan, L., Vognsen, J., Jabre, M. (2020). COVID-19 pandemic prompts the development of a Web-OSCE using Zoom teleconferencing to resume medical students' clinical skills training at Weill Cornell Medicine-Qatar. *BMJ Simulation & Technology Enhanced Learning*, *6*(6), 376–377. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7239650/

Manuel, V. (2010). Sobre la formación de competencias en el sociólogo. *Revista Colombiana de Sociología*, *33*(1), 69–85.

Martínez-González, A., Sánchez-Mendiola, M., Méndez-Ramírez, I., Trejo-Mejía, J. A. (2016). Grado de competencia clínica de siete generaciones de estudiantes al término del internado médico de pregrado. *Gaceta Médica de México*, *152*(5), 679–687.

Martínez-González, A., Sánchez-Mendiola, M., Olivares-Olivares, S. L., Grimaldo-Avilés, J. I., Trejo-Mejía, J. A., Martínez-Franco, A. I., Furman, G. (2020). Colaboración de tres escuelas de medicina de México en un examen clínico objetivo estructurado (ECOE). *Investigación En Educación Médica*, *9*(36), 58–69. https://doi.org/10.22201/FM.20075057E.2020.36.20258

Martínez, A., Soto, G. (2021). *Evaluación para el aprendizaje: enfoque cualitativo.* [MOOC] https://www.coursera.org/lecture/evaluacion-cualitativa/ecoe-KrPQj

Martínez-González, A., Trejo-Mejía, J. A. (2018). ¿Cómo realizar un ECOE? https://doi.org/10.22201/facmed.20075057e.2018.28.18123

Neufeld, V., Norman, G. R. (1985). *Assessing clinical competence*. New York, EUA: Springer Pub. Co.

Prettyman, A. V., Knight, E. P., Allison, T. E. (2018). Objective Structured Clinical Examination From Virtually Anywhere! *The Journal for Nurse Practitioners*, *14*(8), e157–e163. https://doi.org/10.1016/J.NURPRA.2018.05.007

Rivero-López, C. A., Vega-Rodríguez, M. F., Yap-Campos, K., Jiménez-Galván, I., Ponce-Rosas, R. E., Martínez-González, A. (2021). La evaluación de la competencia clínica a través de un Web-ECOE: una experiencia de aplicación. *Investigación En Educación Médica*, *10*(38), 68–75. Recuperado de https://orcid.org/0000-0001-5628-5488 https://orcid.org/0000-0001-9881-2872§https://orcid.org/0000-0002-7603-769X

Trejo-Mejía, J. A., Sánchez-Mendiola, M., Méndez-Ramírez, I., Martínez-González, A. A. (2016). Reliability analysis of the objective structured clinical examination using generalizability theory. *Medical Education Online*, *21*(1). https://doi.org/10.3402/MEO.V21.31650

Trejo-Mejía, J., Blee-Sánchez, G., Peña-Balderas, J. (2014). Elaboración de estaciones para el examen clínico objetivo estructurado (ECOE). *Investigación en Educación Médica*, *3*(9), 56–59. https://doi.org/10.1016/S2007-5057(14)72725-5

Capítulo 35

HABILIDADES SOCIOEMOCIONALES

Laura Delgado Maldonado, Rafael Enrico Sánchez Mayorga

"Todo aprendizaje tiene una base emocional."
PLATÓN

INTRODUCCIÓN

Es bien conocido que las emociones juegan un papel importante en distintos ámbitos de la vida psicosocial de los individuos, tales como la salud y el desempeño académico. En la actualidad se están desarrollando líneas de investigación en la que las habilidades socioemocionales forman parte de los planes y programas de estudio de los alumnos que cursan su educación básica y media superior; esto es, ya no se trata única o exclusivamente que los estudiantes aprendan contenidos de matemáticas, español, ciencias, geografía o historia, solo por mencionar algunas de ellas.

Las más recientes investigaciones en educación apuntan a la necesidad de formar a los estudiantes a ser mejores ciudadanos, con alto sentido de compromiso y solidaridad con su comunidad y responsabilidad con el medio ambiente. En este contexto de educación formal, resulta de particular interés investigar la relación entre la formación de contenidos de los estudiantes y la educación emocional, mirando la habilidades sociales y emocionales como elementos clave que le permitan a los estudiantes alcanzar sus metas académicas, así como el desarrollo de estrategias y habilidades socioemocionales para su futuro personal y profesional.

Las habilidades socioemocionales, como quedará esbozado en el presente artículo, son un campo de investigación ampliamente discutido y analizado por la comunidad científica, y en franca expansión. Sin embargo, bien podríamos acotar, con fines introductorios, que se trata de procesos a través de los cuales las personas ponen en juego diversas estrategias que les permiten gestionar de una manera eficiente sus emociones y toma de decisiones, considerando su experiencia, información de contexto y los distintos recursos cognitivos que pone en juego y en pro del logro de metas, objetivos o simplemente generar un estado de bienestar emocional.

Las estrategias o habilidades se aprenden desde la infancia y a lo largo de todo el ciclo vital, forma parte inherente de los distintos aprendizajes del ser humano y, al menos en teoría, este despliegue de habilidades se va enriqueciendo conforme se crece y madura, de tal suerte que conforme alcanzamos la adultez nos volvemos más capaces para amplificar nuestra capacidad de respuesta y adaptación a las distintas circunstancias, en pro de dar la mejor respuesta posible ante determinado evento, y ser más eficaces conforme las demandas del entorno, los grupos sociales, culturales, etcétera.

Sin embargo, es en la niñez, adolescencia y juventud plena, en que se lleva a cabo este despliegue de desarrollo de habilidades, mismas que se ven consolidadas en la edad adulta. Bajo esta perspectiva, enseñar a los niños, adolescentes y jóvenes acerca de las emociones y, específicamente de la gestión de sus emociones y desarrollo de habilidades sociales, puede contribuir positivamente al logro de sus metas, tanto académicas como de su vida afectiva. Sin duda, esto representa un gran reto para la educación formal e informal, intra o extramuros, que los beneficiará y contribuirá positivamente a lo largo de su vida adulta y en las distintas aristas de su bienestar personal, salud, académico o laboral, solo por mencionar algunos.

LA RELEVANCIA DE LAS HABILIDADES SOCIALES Y EMOCIONALES

En los últimos años, en la elaboración de los programas educativo se ha mostrado un gran interés por el desarrollo de las habilidades socioemocionales, por ello, en las escuelas los docentes son conscientes de la importancia que estas tienen en el desarrollo cognitivo de los estudiantes y su estrecho vínculo con los resultados académicos (Durlak, Weissberg, Dymnicki, Taylor y Schellinger, 2011). Se ha demostrado en estudios que mejorar las competencias sociales como la capacidad de percibir, gestionar o comprender las emociones, mejora los resultados en rendimiento académico, incrementa las conductas positivas y genera mayores niveles de autoestima de los estudiantes (Mira-Galvañ y Gilar-Corbi, 2020).

En este punto, podríamos hacernos dos preguntas, ¿qué son las habilidades socioemocionales? y ¿por qué en los últimos años se ha observado un aumento considerable del interés de los investigadores educativos sobre el tema y también su inclusión en los programas de evaluación?

En términos generales, es un concepto discutido porque cubre una amplia gama de características individuales, como rasgos de personalidad, motivación, preferencias y valores. Sin embargo, las habilidades socioemocionales se consideran un subconjunto de las habilidades, atributos y características que son importantes para el éxito personal y el funcionamiento social. Abarcan disposiciones conductuales, estados internos, orientaciones y gestión de tareas, control de la conducta, así como los sentimientos. También implican las creencias sobre uno mismo y el mundo que caracterizan las relaciones de un individuo con los demás. Para Goodman, Joshi, Nasim y Tyler, (2015) las habilidades socioemocionales son características psicológicas, involucradas en la motivación, el autocontrol y la relación con los demás, que permiten predecir o anticipar una serie de conductas o resultados en la edad adulta.

De Fruyt y Wille, (2015) proporcionaron una definición más conceptual y operativa de las habilidades socioemocionales, sugiriendo que éstas son características individuales que:

1) se originan en la interacción recíproca entre las predisposiciones biológicas y los factores ambientales;
2) se manifiestan en patrones consistentes de pensamientos, sentimientos y comportamientos;
3) continúan desarrollándose a través de experiencias de aprendizaje formales e informales;
4) influyen en resultados socioeconómicos importantes a lo largo de la vida del individuo.

El aprendizaje socioemocional es el proceso a través del cual niños y adultos adquieren y aplican de manera eficaz los conocimientos, actitudes y habilidades necesarias para comprender y manejar las emociones; establecer y lograr metas positivas; sentir y mostrar empatía por y hacia los demás; establecer y mantener relaciones positivas; y tomar decisiones responsables (CASEL, 2015). Estas han ganado una atención cada vez mayor debido a su relevancia para predecir una amplia serie de resultados constantes, que van desde el logro educativo, el comportamiento ciudadano y la empleabilidad, hasta la salud y el bienestar físico y mental (OCDE, 2015).

La importancia que han tenido las habilidades sociales y emocionales es que juegan un papel trascendental en el desarrollo de niños y adolescentes y, combinadas con el rendimiento académico y las habilidades cognitivas, representan un conjunto holístico de habilidades esenciales para el éxito escolar y en la vida futura.

La promoción de estas habilidades en contextos educativos está impulsada por la importancia de mejorar el funcionamiento adaptativo en niños y jóvenes a través del desarrollo de fortalezas personales que pueden servir como herramientas para fomentar resultados positivos y prevenir los negativos (Taylor, Oberle, Durlak y Weissberg, 2017).

Sin embargo, existe un importante número de términos generales alternos para denom-minar las habilidades socioemocionales, como “habilidades no cognitivas”, “habilidades del siglo XXI”, “habilidades de carácter”, “habilidades transferibles” o “habilidades blandas” aunque todos los términos se refieren al mismo espacio conceptual y comparten atributos similares con respecto a su independencia de las habilidades cognitivas, varios autores han concluido que existe una gran variabilidad en el número y la naturaleza de las habilidades incluidas en diferentes modelos y marcos, y que no hay consenso sobre el número y naturaleza de los constructos necesarios para representar las características socioemocionales. La falta de definiciones y marcos conceptuales compartidos presenta un primer desafío y tiene implicaciones para la medición (Abrahams et al., 2019).

Existen diferentes clasificaciones de habilidades socioemocionales, algunas implican pequeños grupos de habilidades mientras que otras integran una mayor diversidad de éstas. Un grupo investigadores encabezado por Elias propuso seis dominios principales de aprendizaje social y emocional, que incluyen: reconocer y manejar las emociones; establecer y lograr metas positivas; apreciar las perspectivas de los demás; establecer y mantener

relaciones positivas; tomar decisiones responsables; y manejar situaciones interpersonales de manera constructiva (Elias et al., 1997).

Los modelos para las habilidades socioemocionales involucran principalmente características de personalidad. Como en el proyecto desarrollado por Collaborative for Academic, Social, and Emotional Learning (CASEL, 2015). Programas efectivos de aprendizaje social y emocional en Chicago propone un modelo con cinco variables centrales: autoconocimiento, autogestión, conciencia social, habilidades de relación y toma responsable de decisiones.

Asimismo, un estudio internacional sobre Habilidades Sociales y Emocionales de la OCDE, (2017) evalúa a estudiantes de 10 y 15 años en varias ciudades y países del mundo. El estudio se basó en un marco conocido en el campo de las habilidades sociales y emocionaales como Big Five (cinco grandes), mismo que proporciona un esquema general (ver Figura 1) de cómo deben organizarse estas habilidades. El modelo se organiza jerárquicamente las cinco categorías generales de habilidades que se pueden a su vez subdividir en habilidades más condensadas y que pueden considerarse de orden inferior.

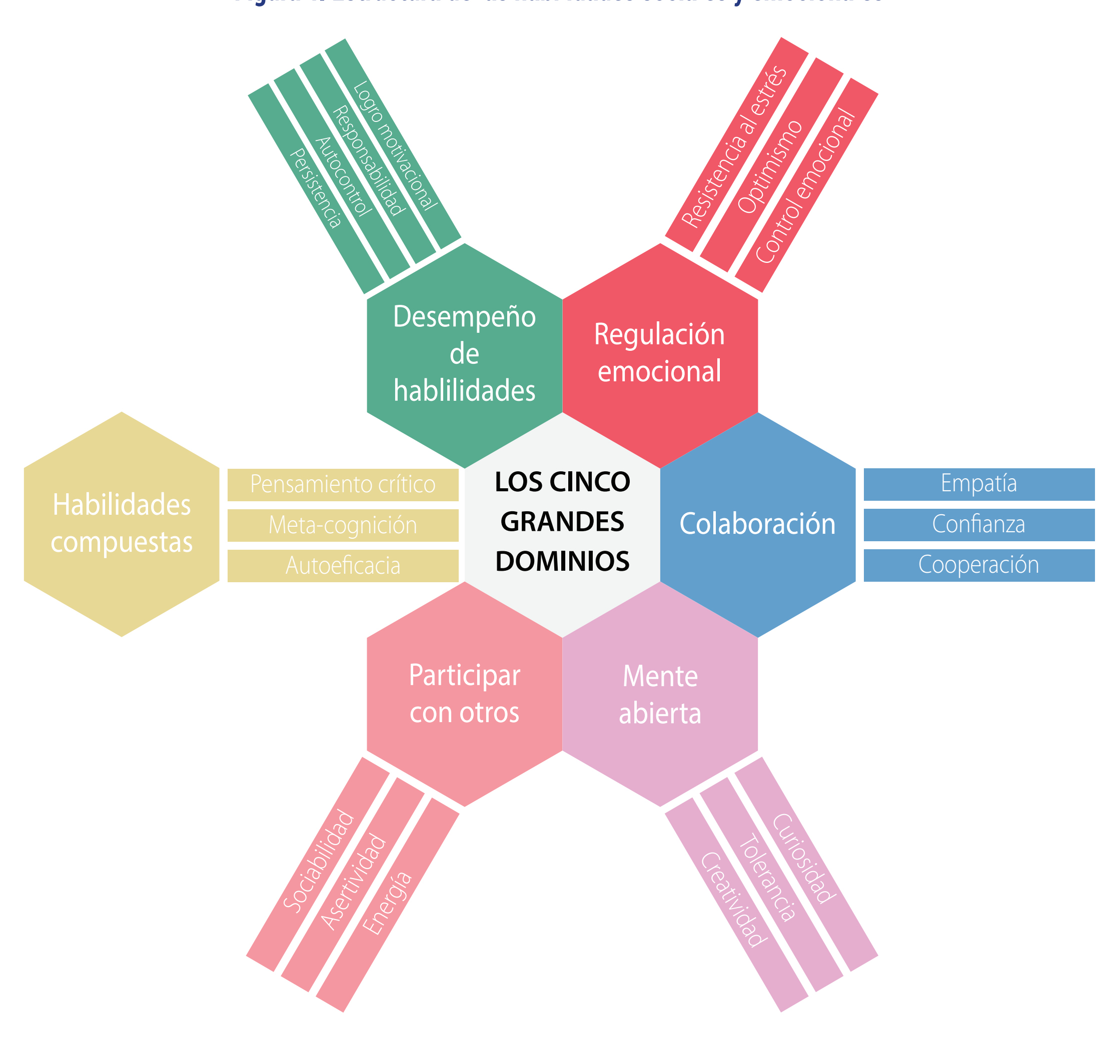
Además de mostrar su similitud mutua, estas agrupaciones también aseguran una consideración sistemática, integral y equilibrada de las habilidades sociales y emocionales de los individuos.

De acuerdo con este estudio, la estabilidad emocional parece ser la más relevante de las cinco grandes dimensiones para la satisfacción con la vida. También sugiere una fortaleza notablemente similar de las relaciones para dos resultados de calidad de vida: la vida y la satisfacción laboral. Establecer la causa exacta de estas relaciones es desafiante, sin embargo, también se ha encontrado que el bienestar promueve cambios positivos de personalidad (entre otros factores) que pueden mejorar simultáneamente las habilidades sociales y emocionales de las personas y su calidad de vida.

Por otra parte, Durlak, Weisberg, Dymnicki, Taylor y Schellinger (2011) defienden que los programas de aprendizaje socioemocional deberían fomentar cinco conjuntos amplios de competencias: autoconocimiento, autogestión, conciencia social, habilidades de relación y toma responsable de decisiones.

El modelo elaborado por Primi con otros autores presenta cinco amplios dominios de habilidades socioemocionales: Autogestión la cual se relaciona con el Conocimiento de la Literatura de Personalidad; las dimensiones interpersonales de Compromiso con los Demás conceptualmente relacionado con Extraversión; Amistad relacionada con Amabilidad; Regulación de Emociones Negativas asociada a Afectividad Negativa o Neuroticismo; y Mentalidad Abierta asociada a Apertura a la Experiencia. Estos dominios y facetas socioemocionales pueden servir como un marco integrador que se puede usar para construir evidencia, transversal y longitudinalmente, sobre los resultados consecuentes de las habilidades socioemocionales. Habilidades adquiridas a lo largo del desarrollo de los niños y jóvenes (Primi et al., 2017; Abrahams, et al., 2019).

Figura 1. Estructura de las habilidades sociales y emocionales



Traducido de OECD, (2018). Social and Emotional Skills Well-being, connectedness and success. Paris, France: OECD Publishing

Tabla 1. Dominios y facetas de las habilidades socioemocionales (Primi et al., 2017)

Dominio	Faceta	Definición
Autogestión	**Determinación**	Establecer metas y altos estándares, motivarse, trabajar muy duro y aplicar uno mismo completamente a la tarea, trabajo o proyecto en cuestión.
	Organización	Poseer habilidades de organización y una atención meticulosa a los detalles que sean útiles para planificar y ejecutar planes para alcanzar objetivos a más largo plazo.
Interactuar con los demás	**Centro de atención**	Centrar la atención y concentrarse en la tarea actual, y evitar distracciones.
	Persistencia	Superar obstáculos para alcanzar metas importantes.
	Responsabilidad	Poseer habilidades de gestión del tiempo, ser puntual y honrar los compromisos.
	Iniciativa social	Acercarse y conectar con otros, tanto amigos como extraños, iniciando, manteniendo y disfrutando del contacto social y las conexiones.
	Asertividad	Hablar, expresar opiniones, necesidades y sentimientos, y ejercer influencia social.
	Entusiasmo	Mostrar pasión y entusiasmo por la vida; abordar las tareas diarias con energía, emoción, y una actitud positiva.
Amistad	**Compasión**	Usar la empatía y las habilidades de toma de perspectiva para comprender las necesidades y los sentimientos de los demás, actuando sobre esa comprensión con amabilidad y consideración.
	Respeto	Tratar a los demás con respeto y cortesía.
	Confianza	Asumir que los demás, en general, tienen buenas intenciones y perdonar a los que han obrado mal.
Regulación de emociones negativas	**Modulación del estrés**	Modulación de la ansiedad y respuesta al estrés.
	Auto confianza	Sentirse satisfecho con uno mismo y con la vida actual, tener pensamientos positivos sobre uno mismo y mantener expectativas optimistas.
	Tolerancia a la frustración	Regular el temperamento, la ira y la irritación; manteniendo la tranquilidad y la ecuanimidad ante las frustraciones.
Mentalidad abierta	**Curiosidad intelectual**	Mostrar interés por las ideas y pasión por el aprendizaje, la comprensión y la exploración intelectual.
	Creatividad	Generar formas novedosas de pensar o hacer las cosas a través de la experimentación, el retoque, el aprendizaje del fracaso, la perspicacia y la visión.
	Imaginación e interés artístico	Valorar, apreciar y disfrutar el diseño, el arte y la belleza, que pueden experimentarse o expresarse en la escritura, las artes visuales y escénicas, la música y otras formas de autorrealización.

La evaluación de habilidades socioemocionales se considera un desafío debido a múltiples problemas de medición como: estilos de respuesta, problemas de consistencia y falta de convergencia, además de otras medidas que también son visibles en el campo de la evaluación de las diferencias individuales. El desarrollo y mejora de herramientas y métodos de evaluación socioemocional es, por tanto, una prioridad.

MÉTODOS DE EVALUACIÓN PARA LAS HABILIDADES SOCIOEMOCIONALES

Una deficiencia de las escalas de autoevaluación para medir las habilidades socioemocionales corresponde a la dependencia de los reactivos tipo Likert, debido a que, por lo general, en este tipo de reactivos se presenta a los estudiantes afirmaciones a las que responden marcando un punto en una escala, por ejemplo: desacuerdo/de acuerdo. Si bien se consideran convenientes, eficientes y apropiados para pruebas de bajo riesgo, tienen una variedad de sesgo de respuesta que pueden afectar la confiabilidad y validez de las puntuaciones obtenidas (Abrahams et al., 2019).

Para los desafíos de la evaluación y las innovaciones en la investigación de habilidades socioemocionales, Abrahams, y sus colaboradores en 2019 sugirieron varios métodos alternativos para medirlas, que complementan las herramientas basadas en escala de autoinforme. Asimismo, realizaron una revisión de medidas alternativas que pueden corregir las deficiencias de este tipo de escalas, describiendo una serie de problemas potenciales asociados con el uso de las escalas tipo Likert.

A continuación, se muestran algunos de los principales métodos de evaluación para las habilidades socioemocionales.

Corrección por aquiescencia

La aquiescencia se define como un sesgo de respuesta que se caracteriza por ser una tendencia general de un individuo a estar consistentemente de acuerdo (decir sí) o en desacuerdo (decir no) con los reactivos del cuestionario, independientemente de su contenido, y sus efectos.

Una solución que ayuda a contrarrestar la aquiescencia es incluir tanto reactivos positivos en la dirección del constructo a medir como negativos en la dirección opuesta en las escalas. Equilibrando aproximadamente, el mismo número de reactivos positivos y negativos.

Hay varios métodos a posteriori que corrigen la aquiescencia usando procedimientos estadísticos. La mayoría de estos procedimientos se basan en escalas completamente balanceadas, sin embargo, también permiten corregir en escalas cuasi balanceadas (mismo número de ítems positivos y negativos, pero con saturaciones desiguales) y balanzas desequilibradas (diferente número de ítems positivos y negativos).

Los métodos estadísticos basados en el modelo de factor común para controlar el sesgo de aquiescencia son: el análisis factorial exploratorio (AFE) el cual propone evaluar la dimensionalidad y la estructura de una escala de personalidad equilibrada teniendo en cuenta los efectos potenciales de la respuesta aquiescente, y corregir las estimaciones de rasgos

individuales para la aquiescencia. (Ferrando et al., 2003), el modelo de análisis factorial de intercepción aleatoria (RIFA) asume que los coeficientes lineales (intersecciones y cargas factoriales) que vinculan las variables observadas con los factores latentes son coeficientes fijos es decir, comunes para todos los participantes (Maydeu-Olivares y Coffman, 2006) y el método ipsativo propuesto por Chan y Bentler (1993) donde los datos se denominan ipsativos si están sujetos a una restricción de suma constante para cada observación, usualmente, los datos ipsativos son la consecuencia de la transformación de sus correspondientes datos preipsativos.

Viñetas de anclaje

La técnica de anclar viñetas se introdujo para corregir el sesgo de medición, en particular, el llamado efecto de grupo de referencia, en las respuestas de tipo Likert (King, Murray, Salomon y Tandon, 2004). Las viñetas de anclaje son textos breves que representan a un individuo en un escenario hipotético que manifiesta el rasgo de interés en mayor o menor grado, los encuestados usan el mismo formato de respuesta y la misma escala de autoevaluación de su propio comportamiento o rasgos. Por lo general, los encuestados califican varias viñetas, que representan varios niveles del rasgo.

Los principales supuestos de medición del método de viñetas de anclaje son la consistencia de la respuesta la cual significa que los encuestados usan las categorías de respuesta de la misma manera cuando califican las viñetas que cuando se califican a sí mismos en lugar de imponerse estándares más altos o más bajos que los personajes de las viñetas, y la equivalencia de viñetas significa que todos los encuestados perciben que una viñeta representa el mismo concepto subyacente, con viñetas en una serie, todas vistas como parte de una escala unidimensional (King et al., 2004).

Las autoevaluaciones de las personas se pueden volver a escalar en función de sus calificaciones en las viñetas de anclaje; en las evaluaciones de PISA 2012, este cambio de escala condujo a la eliminación de la paradoja motivación-logro. Asimismo, las viñetas se diseñaron para garantizar la unidimensionalidad y un poder discriminatorio considerable en los niveles de rasgos (OCDE, 2013).

Métodos múltiples informantes

El método basado en múltiples informantes, es el grado en que dos evaluadores independientes están de acuerdo en la evaluación del comportamiento de los estudiantes, como padres, amigos y maestros, se consideran comúnmente como recursos para evaluar las habilidades emocionales y del comportamiento, la adopción de esta evaluación, aunque lleva mucho tiempo, también puede ayudar a corregir los problemas asociados con el uso de escalas tipo Likert y aportar información adicional en la evaluación de las habilidades socioemocionales.

Algunos investigadores establecen que las habilidades socioemocionales deben basarse en calificaciones de múltiples informantes, en lugar de una sola fuente (John y De Fruyt, 2015). Sin embargo, al incluir múltiples informantes para evaluar las habilidades socioemocionales, también se deben tener en cuenta el tipo de informante, ya que puede haber una

variabilidad sustancial entre las percepciones estos. Las calificaciones de diferentes informantes comparten cierta variación, pero también brindan perspectivas únicas sobre la posición y el desarrollo de habilidades de los niños y adolescentes.

Tales discrepancias entre los informantes no son sorprendentes, ya que las habilidades socioemocionales son construcciones multidimensionales que se manifiestan de diversas formas en diferentes contextos (Abrahams et al., 2019). Por ejemplo, los profesores se pueden basar en un marco de referencia mucho más amplio al comparar a un alumno con los otros alumnos a los que enseñan, para caracterizar las habilidades de sus alumnos. Por el contrario, el alcance de los padres suele ser mucho más estrecho e idiosincrásico (ver John y De Fruyt, 2015).

Escalas compuestas

Las escalas compuestas generalmente se refieren a descripciones amplias que a menudo que refieren a construcciones multidimensionales que aprovechan múltiples habilidades o rasgos, por ejemplo, la "ciudadanía global", "espíritu empresarial" o "liderazgo". El desafío clave para la evaluación de tales términos de habilidades híbridas y multidimensionales es primero descomponerlos en sus habilidades constitutivas definidas en una clasificación de habilidades existente y, posteriormente, pensar en formas de recombinar psicométricamente de los puntajes de habilidades que constituyen para que sirvan como un indicador de la habilidad híbrida (Abrahams et al., 2019).

En la literatura sobre los efectos de criterio de los rasgos de personalidad y la validez de criterio de escalas compuestas específicos subrayan la necesidad de claridad sobre las estructuras de escalas compuestas. A nivel conceptual, la construcción de teorías busca comprender cómo los rasgos de personalidad afectan los resultados y para qué criterios los rasgos específicos son más o menos destacados. A nivel de medición, la investigación de validación se ocupa de probar los efectos de criterio de instrumentos específicos (por ejemplo, Hough y Ock, 2013). Con respecto a estos dos aspectos, los investigadores podrían conceptualizar y medir los rasgos de personalidad en varios niveles de amplitud versus fidelidad (por ejemplo, faceta, orden superior y nivel compuesto), lo que complica aún más la imagen. La comprensión estructural de las escalas compuestas es importante para examinar todos estos problemas.

Evaluaciones de elección forzada

El formato de elección forzada es una técnica de construcción de pruebas, en esta se pueden presentan dos o más declaraciones de manera comparativa, se les pide a los individuos que consideren múltiples afirmaciones a la vez y que califiquen en qué medida los reactivos se aplican a ellos en relación con los demás reactivos. Es decir, se requiere que los encuestados clasifiquen los reactivos o seleccionen dos elementos que sean "más parecidos a mí" y "menos parecidos a mí", respectivamente.

En una prueba de elección forzada en la que un bloque se compone de dos afirmaciones emparejadas, si se utiliza el método de puntuación tradicional, la puntuación de un

encuestado en una dimensión es igual al número de afirmaciones que elige para medir esa dimensión. El método de puntuación es similar cuando cada bloque consta de más de dos reactivos. Así, para cualquier individuo, la suma de sus puntuaciones es un valor fijo, que genera datos ipsativos.

Para resolver fundamentalmente los problemas de los datos ipsativos, los científicos psicométricos han construido algunos modelos de teoría de respuesta al ítem (IRT, por sus siglas en inglés) en los últimos años. Uno de ellos es el modelo TRI thurstoniano de Brown y Maydeu-Olivares (2011). El modelo puede estimar la puntuación real del encuestado directamente a través de su patrón de respuesta. Por lo tanto, los investigadores pueden analizar los puntajes normativos y estimar la confiabilidad de la prueba utilizando la función de información de la prueba (Xiao, Yue, Liu, Hongyun y Li, Hui, 2017)

Rúbricas como anclas de escala de Likert

Para Abrahams et al. (2019) una alternativa es usar escalas ancladas en el comportamiento, usando los niveles de habilidad definidos en las rúbricas como anclas. Las rúbricas son un conjunto de criterios de calidad para calificar un determinado tipo de desempeño y se han utilizado con frecuencia en la evaluación formativa para brindar retroalimentación a los estudiantes y en la evaluación sumativa para calificar el trabajo de los estudiantes tanto con calificaciones propias y de pares.

Los criterios de evaluación formulados en las rúbricas brindan ventajas sobre otros métodos de evaluación tanto en términos de la capacidad de autoevaluación de los estudiantes como en términos de su capacidad para regular su desempeño. En este sentido, las rúbricas pueden ser más efectivas que las simples evaluaciones de competencias basadas en una escala de Likert (Bartlett et al., 2015).

Diferencias intraindividuales y desempeño dinámico

La evaluación ambulatoria recopila datos en la vida cotidiana de la persona examinada varias veces al día, los individuos informan sobre sus sentimientos, pensamientos o comportamientos actuales tal como lo experimentan en su entorno cotidiano. Aunque se tiene varios tipos y nombres como métodos diarios, evaluación momentánea, métodos de experiencia cotidiana, muestreo de experiencia.

La mayoría de los métodos de evaluación que se utilizan actualmente en la medición de habilidades socioemocionales están diseñados para medir las diferencias entre los estudiantes. Las evaluaciones ambulatorias, por el contrario, son adecuadas para la evaluación de la variabilidad intraindividual, es decir, cómo la posición del estudiante en las habilidades socioemocionales fluctúa a través de diversas situaciones. Los puntos múltiples de datos intraindividual brindan una diversidad de información sobre cómo las experiencias pueden cambiar en contextos y marcos de tiempo cortos lo que brinda información sobre la dinámica del funcionamiento del estudiante.

Biodatos y pruebas de juicio situacional

Los estudios de biodatos emplean preguntas estandarizadas en los individuos sobre detalles biográficos, como la frecuencia de ocurrencia de ciertos eventos, comportamientos o experiencias en el pasado. Dado que las medidas de biodatos son menos propensas al engaño que los inventarios de autoinforme tradicionales y consumen menos tiempo que la evaluación ambulatoria, podrían ser un método práctico para la evaluación de comportamientos específicos relacionados con habilidades.

La información de biodatos se ha utilizado principalmente para las admisiones universitarias y ha demostrado una validez incremental para la predicción del rendimiento de los estudiantes universitarios, la personalidad de los Cinco Grandes, así como para evaluar rasgos de personalidad específicos.

Las medidas de biodatos proporcionan un método estructurado y sistemático para recopilar y calificar información sobre los antecedentes y la experiencia de un individuo en cambio el inventario de medida y juicio situacional (SJI, por sus siglas en inglés) son pruebas de opción múltiple destinadas a evaluar cómo un solicitante (para un trabajo o, en este caso, para la universidad) podría reaccionar en diferentes contextos relevantes (Motowidlo y Tippins, 1993). En el contexto de la universidad, ambas medidas tienen el potencial de aumentar la validez relacionada con el criterio sobre las medidas tradicionales, porque el contenido de los ítems se puede adaptar a dimensiones específicas del desempeño de los estudiantes en la universidad y a las metas de una universidad en particular o del proceso de admisión a la universidad.

En las pruebas de juicio situacional a los encuestados se les proporciona un conjunto de escenarios hipotéticos acompañados de varios cursos de acción plausibles. Según el diseño se les pide a los examinados que clasifiquen o califiquen las diferentes respuestas según su idoneidad en esa situación. Alternativamente, a los participantes se les pueden ofrecer múltiples opciones de respuesta y se les puede pedir que seleccionen la respuesta que define su comportamiento en dicha situación.

Los biodatos y los SJI también pueden tener una mayor utilidad práctica sobre las evaluaciones subjetivas alternativas, como ensayos o cartas de referencia que se usan comúnmente en las admisiones universitarias, porque estas brindan un método justo y estandarizado para obtener y calificar información sobre la amplia gama de antecedentes educativos y experiencias sociales que los solicitantes puedan haber tenido.

Medidas de desempeño

En las medidas de desempeño específicas de habilidades, principalmente para evaluar el pensamiento crítico y la creatividad, podría implicar que estas dos habilidades parecen prestarse mejor para las pruebas de rendimiento que otras. Sin embargo, el objetivo es integrar múltiples habilidades en una sola prueba. Por ejemplo, la evaluación de preparación para la universidad y el trabajo involucra no solo pensamiento crítico sino también razonamiento analítico, resolución de problemas y habilidades de comunicación escrita (Partnership for 21st Century Skills, 2009).

En la literatura, cada vez existe más evidencia empírica que nos lleva a pensar que fomentar las competencias socioemocionales de los estudiantes "puede ser una herramienta útil para abordar las emociones negativas incipientes en diferentes contextos sociales o de aprendizaje y, en última instancia, ser un factor que facilite de manera efectiva la resolución de problemas y el procesamiento de la información" (Cassullo y García, 2015, p. 215).

Los resultados de una investigación a partir de una metodología a nivel universitario como parte de proyectos de investigación e innovación relacionados con la docencia. López-Mondéjar y Tomás, (2016) mostraron que el uso de la metodología del aprendizaje cooperativo en el ámbito universitario favorece el desarrollo en los estudiantes de habilidades y competencias socioemocionales como la empatía, el asertividad y el consenso, con la capacidad de ponerse en el lugar de los demás, o empatía. La importancia de este tipo de investigación es de total relevancia ya que las universidades preparan a los ciudadanos que en el futuro participarán activamente en la sociedad, por lo que el desarrollo de estas competencias socioemocionales en los jóvenes cobra especial relevancia.

Por otra parte, para Schleicher (2018) afirma que el mayor dilema de la comunidad educativa es que las habilidades cognitivas, son las más fáciles de enseñar y probar, son las que, a su vez, "también son más fáciles de digitalizar, automatizar y tercerizar", la escolarización de hoy debe ser mucho más acerca de las formas de pensar (que involucran la creatividad, el pensamiento crítico, la resolución de problemas y el juicio), las formas de trabajar (incluidas la comunicación y la colaboración), las herramientas para trabajar (incluida la capacidad de reconocer y explotar el potencial de las nuevas tecnologías) y sobre la capacidad de vivir en un mundo multifacético como ciudadanos activos y responsables (p. 31).

En las investigaciones se han explorado factores que contribuyen al desempeño exitoso de los estudiantes como son: socioculturales, familiares e individuales (Hattie, 2009). Asimismo, otras investigaciones a partir de teoría cognitiva social ven el funcionamiento humano de una manera transaccional, dependiendo de las interacciones recíprocas entre los comportamientos de un individuo, sus factores personales internos (por ejemplo, pensamientos y creencias) y eventos ambientales (Bandura, 1997).

A MANERA DE CIERRE

En la evidencia de múltiples investigaciones, es cada vez más contundente confirmar datos consistentes reportados en la literatura internacional que afirman sobre el impacto que tienen las habilidades socioemocionales en el rendimiento académico; lo que hace imprescindible considerar que en el diseño de las políticas públicas en educación se comience a pensar, de manera clara y explícita en la necesidad de fortalecer la dimensión social y emocional de los estudiantes al igual que de los profesores y en los contextos en los cuales se desarrollan. En este sentido, lo presentado da aún mayor relevancia a la necesidad de integrar en los programas de formación docente contenidos relacionados con el aprendizaje socioemocional de sus estudiantes, como también ofrecer un entrenamiento práctico sobre cómo implementar dinámicas interpersonales en los salones de clases.

Las habilidades socioemocionales a lo largo de la vida tienen gran relevancia para predecir una amplia serie de resultados constantes, que van desde el logro educativo, el comportamiento ciudadano y la empleabilidad, hasta la salud y el bienestar físico y mental. Por ello, la educación en todos sus niveles debería posibilitar que los alumnos, como futuros ciudadanos, construyan las habilidades que les ayuden a reflexionar y desarrollar habilidades emocionales y sociales que faciliten la convivencia entre los sujetos en diversos escenarios en los que puedan demostrar valores y actitudes tales como el respeto, la empatía, la participación y la comunicación en pro de una sociedad equitativa que brinde herramientas para el bienestar individual y social.

Los desafíos de la evaluación y las innovaciones en la investigación de habilidades socioemocionales llevarán a los investigadores en educación a revisar y aplicar métodos alternativos para medirlas, que complementen las herramientas basadas en escala de autoinforme.

REFERENCIAS

Abrahams, L., Pancorbo, G., Primi, R., Santos, D., Kyllonen, P., John, O. P., De Fruyt, F. (2019). Social-emotional skill assessment in children and adolescents: Advances and challenges in personality, clinical, and educational contexts. *Psychological assessment*, *31*(4), 460–473. https://doi.org/10.1037/pas0000591

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York, NY: W.H. Freemanand Company.

Becker, E. S., Goetz, T., Morger, V., Ranellucci, J. (2014). The importance of teachers' emotions and instructional behavior for their students' emotions—An experience sampling analysis. *Teaching and Teacher Education*, *43*, 15–26. http://dx.doi.org/10.1016/j.tate.2014.05.002

Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement*, 71, 460–502.

Casullo, G.L. y García, L. (2015). Estudio de las Competencias Socio Emocionales y su Relación con el Afrontamiento en Futuros Profesores de Nivel Medio. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, *18*(1), 213-228.

Chan, W., Bentler, P. M. (1993). The Covariance Structure Analysis of Ipsative Data. *Sociological Methods & Research*, *22*(2), 214-247. https://doi.org/10.1177/0049124193022002003

Collaborative for Academic, Social and Emotional Learning (Colaboración para el Aprendizaje Académico, Social y Emocional). CASEL.org

Durlak J. A., Weissberg R. P., Dymnicki A. B., Taylor R. D., Schellinger K. B. (2011). The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Development.* https://doi.org/10.1111/j.8624.2010.01564-1467.x

De Fruyt, F., Wille, B., John, O. P. (2015). Employability in the 21st century: Complex (interactive) problem solving and other essential skills. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 8(2)*, 276–281. http://dx.doi.org/10.1017/iop.2015.33

Ferrando, P. J., Lorenzo-Seva, U., Chico, E. (2003). Unrestricted Factor Analytic Procedures for Assessing Acquiescent Responding in Balanced, Theoretically Unidimensional Personality Scales. *Multivariate Behavioral Research*, *38*(3), 353-374

Goodman, A., Joshi, H., Nasim, B., Tyler, C. (2015). Social and emotional skills in childhood and their long-term effects on adult life. Recuperado de https://www.eif.org.uk/report/social-and-emotional-skills-in-childhood-and-their-long-term-effects-on-adult-life

Hough LM, Johnson JW. 2013. Use and importance of personality variables in work settings. En Handbook of Psychology, Vol. 12: Industrial and Organizational Psychology, ed. IB Weiner, N Schmitt, S Highhouse, pp. 211–43. New York: Wiley.

John, O. P., De Fruyt, F. (2015). Education and social progress: Framework for the longitudinal study of social and emotional skills in cities. Paris: OECD Publishing.

King Gary, Murray Christopher JL, Salomon Joshua A, Tandon Ajay. Enhancing the Validity and Cross-Cultural Comparability of Survey Research. *American Political Science Review*. 2004 Feb; *98*(1):191–207.

Maydeu-Olivares, A.,Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods, 11*(4), 344–362. https://doi.org/10.1037/1082-989X.11.4.344

Mira-Galvañ, M. J., Gilar-Corbi, R. (2020). Design, Implementation and Evaluation of an Emotional Education Program: Effects on Academic Performance. *Frontiers in psychology*, *11*, 1100.

Motowidlo, S. J., Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology*, *66*(4), 337–344.

López-Mondéjar L. M., Tomás L. M. (2017) Development of socio-emotional skills through cooperative learning in a university environment. *Procedia - Social and Behavioral Sciences* 237 432 – 437

Organisation for Economic Cooperation and Development (OECD). (2015). *Skills for social progress: The power of social and emotional skills*. Paris, France: OECD Publishing.

Organization for Economic Cooperation and Development (OECD). (2018). Social and Emotional Skills Well-being, connectedness and success. Paris, France: OECD Publishing.

Partnership for 21st Century Skills. (2009). Assessment: A 21st Century Skills Implementation Guide. Recuperado de http://www.p21.org/ storage/documents/p21-stateimp_assessment.pdf

Primi, R., John, O. P., Santos, D., & De Fruyt, F. (2017). SENNA inventory. São Paulo, Brazil: Institute Ayrton Senna.

Schleicher, A. (2018). World Class: How to build a 21st-century School System, Strong Performers and Successful Reformers in Education. Paris: OECD Publishing.

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development, 88*, 1156 –1171.

Xiao, Y., Liu, H., Li, H. (2017). Integration of the Forced-Choice Questionnaire and the Likert Scale: A Simulation Study, Frontiers in Psychology. https://doi.org/10.3389/fpsyg.2017.00806

Sección IV

ASPECTOS PRÁCTICOS DE USO Y APLICACIÓN

Capítulo 36

ALINEAMIENTO DEL CURRÍCULO, MÉTODOS DE ENSEÑANZA Y EVALUACIÓN

Nancy Sofía Contreras Michel, Melchor Sánchez Mendiola

"La vida es un currículo único para cada estudiante."
Joyce Rachelle

El presente capítulo tiene el propósito de orientar e inspirar a los docentes a realizar un cambio profundo de perspectiva sobre sus prácticas, no solo de evaluación, sino de enseñanza en sintonía con el currículo. De acuerdo con Bamber, Trowler, Saunders y Knight (2009), el cambio es algo que sucede y la continuidad algo que persiste. En el ámbito educativo la mejora requiere de cambios, y el docente puede ser un agente de cambio cuyo interés está en la mejora de la calidad del aprendizaje, los métodos de enseñanza, la evaluación y el currículo. Sin embargo, en la educación media superior y superior, si bien las situaciones cambian, algunas cosas permanecen igual. Independientemente de la postura que se tome sobre cómo se ejerce la docencia, es importante tener en cuenta que un cambio puede tener efectos tanto positivos como negativos, por lo que es fundamental emplear de manera eficiente los recursos disponibles.

En ese sentido, la mejora implica un cambio de perspectiva deliberado y puede ser tan simple como hacer lo mismo, pero mejor, es decir, mejorar lo que ya existe. Por ejemplo, mejorar los materiales educativos o innovar desde la práctica, renovando enfoques y cuestionando creencias. En este capítulo se desarrollan tres puntos clave a considerar en todo proceso de enseñanza y aprendizaje: a) currículo, b) evaluación y c) métodos de enseñanza, a fin de describir cuál es el papel que juega cada uno en este proceso. Más allá de eso, se considera también la influencia que tiene la interacción entre estos elementos, para que los alumnos tengan una experiencia de aprendizaje fructífera y el docente sea un hábil orquestador de tales experiencias de aprendizaje.

ALINEACIÓN CONSTRUCTIVA DE ENSEÑANZA, CURRÍCULO Y EVALUACIÓN

Diversos autores han propuesto esquemas y modelos para alinear lo que se enseña y cómo se enseña, con lo que se aprende (Asobasi, 2019; Bamber et al., 2009; Biggs y Tang, 2011; Martone y Sireci, 2009; Pendergast y Bahr, 2005). El tema no es menor ya que la palabra "alineación" o "alineamiento" significa varias cosas en el mundo de la educación, desde la definición de diccionario en la que alinear quiere decir colocar en una línea dos o más elementos para su correcto funcionamiento (con sus implicaciones positivistas de linealidad), hasta modelos más complejos y sofisticados que requieren planeación, coordinación y colaboración entre los elementos del sistema (Martone y Sireci, 2009). En el salón de clases, el alineamiento instruccional puede referirse a la concordancia y congruencia entre los objetivos del profesor, las actividades de aprendizaje y los métodos de evaluación, de forma que estos elementos se apoyen mutuamente; a nivel institucional el alineamiento curricular puede tratarse del grado en el que los diferentes planes y programas de estudio están estructurados y se entrelazan en un todo integrado.

Podemos considerar el alineamiento curricular como el grado en el que las expectativas educativas del currículo, las estrategias de enseñanza utilizadas y los métodos de evaluación están de acuerdo y son congruentes, para en conjunto guiar al sistema en el logro del aprendizaje de los estudiantes (Martone y Sireci, 2009; Webb, 1997). Las evaluaciones deben permitir que los estudiantes demuestren sus conocimientos y habilidades con respecto a las metas descritas en el marco curricular, de forma tal que las evaluaciones puedan interpretarse con validez de forma apropiada, y se integren en un esquema que favorezca el aprendizaje utilizando la evaluación del, para y como aprendizaje (capítulo 1). En un mundo ideal, lo que se logre demostrar con las evaluaciones debería derivarse de lo que se espera en el currículo formal y de lo que se enseña por el profesorado, sin embargo, los diferentes tipos de currículo (formal, vivido, oculto, nulo) generan que lo que aprenden los estudiantes no siempre es lo que está en el plan de estudios o lo que pretendemos los docentes.

En este complejo ecosistema educativo, que se comporta como un sistema adaptativo complejo, debemos estar alerta a lo que ocurre durante todo el proceso, con una visión sistémica y utilizando las herramientas de evaluación de manera constante y eficaz, para identificar cuando la falta de este alineamiento ideal se convierte en un problema serio. El mensaje que envían las instituciones educativas con el plan de estudios (y los docentes con sus estrategias de enseñanza y evaluación) sobre lo que es importante para aprender, es captado por los estudiantes quienes actúan en consecuencia. Por ello debemos estar alertas y conscientes de estas señales y las respuestas de los estudiantes, para responder de manera dinámica teniendo en mente el aprendizaje significativo de los estudiantes. En la Figura 1 se esquematizan los componentes de este triángulo interactivo, en el que se representa la sinergia deseable en cualquier práctica relacionada con la evaluación del y para el aprendizaje, el currículo y los métodos de enseñanza.

Figura 1. Los tres elementos para tomar en cuenta en el alineamiento del proceso educativo: currículo, evaluación y métodos de enseñanza



Bajo la premisa de que la evaluación debe estar alineada con el aprendizaje, usando el concepto de "evaluación para el aprendizaje" (capítulo 1), es menester utilizar modelos que privilegien el aprendizaje del estudiantado, como el propuesto por Biggs (2011), del "alineamiento constructivo" (Figura 2). Este alineamiento se rige bajo el principio de que el estudiante construye su propio aprendizaje a través de actividades pertinentes y relevantes, y es tarea del docente crear ambientes que soporten dichas actividades de aprendizaje de forma tal que se logren los resultados deseados (Biggs y Tang, 2011). El elemento clave de este modelo es que todos los componentes del sistema (currículo y metas educativas, métodos de enseñanza y estrategias de evaluación) estén alineados entre ellos, de forma orquestada y en sintonía.

Figura 2. Esquema del alineamiento constructivo del currículo con los métodos de enseñanza y evaluación

Adaptado de Biggs y Tang, 2011.

Para que la evaluación en los espacios educativos tenga impacto en el aprendizaje es importante que se involucre a los distintos grupos de interés: profesores, alumnos y autoridades educativas, para que la incorporen como parte del proceso de enseñanza aprendizaje y no sea un elemento aislado que va al final del curso. La evaluación del y para el aprendizaje es un reto que los docentes enfrentan día con día, por ello es importante que se profesionalicen en este campo, a fin de que desarrollen habilidades que les permitan ofrecer mejores oportunidades de aprendizaje a los alumnos. Cuando tiene lugar la evaluación del y para el aprendizaje en los espacios educativos, es mayor el beneficio para los alumnos que se encuentran en el proceso de aprendizaje si se avanza en la realización de mejores prácticas educativas.

CURRÍCULO

Es evidente la necesidad de contar con un currículo cohesivo, comprensible, que intencionalmente conecte los resultados de aprendizaje, los métodos de enseñanza y la evaluación. En cada ciclo escolar, el docente se enfrenta a la interrogante: ¿qué aprendieron mis alumnos de las experiencias de aprendizaje que tuvimos, en función del currículo? El currículo es una herramienta que le proporciona al docente un marco de organización que le permite dar seguimiento continuo al progreso de los alumnos en su nivel de dominio, en cuanto a alcanzar los resultados de aprendizaje esperados. De tal forma que el diseño curricular debería servir como predictor del éxito académico del alumno.

Esto es lo deseable para que los docentes puedan llevar a cabo su labor, de forma que cuenten con un currículo holístico, cuyos componentes estén alineados, con resultados de aprendizaje claros y precisos, susceptibles de ser evaluados con instrumentos y estrategias de evaluación acordes con esos resultados. Ello permitirá llevar a cabo experiencias de aprendizaje relevantes, con métodos de enseñanza y evaluación que cuenten con evidencia sobre el proceso y los resultados educativos, al tiempo que propicien el aprendizaje. Desafortunadamente con frecuencia encontramos currículos con poca articulación, con listas extensas de contenidos y una mención superficial de las estrategias de evaluación. Independientemente de qué tan estructurado esté el currículo, el docente puede poner en práctica habilidades teórico metodológicas, en función de las condiciones en las que desempeña su trabajo, para realizar sus prácticas educativas con el interés de proporcionar valor agregado a los alumnos, a sabiendas que su intervención puede hacer una gran diferencia en las experiencias de aprendizaje.

Etimológicamente, currículo significa “plan a ejecutar”, con el deseo de llegar a un punto final. Pero, para llegar a un lugar determinado es necesario saber cuál es el punto de partida y de ahí determinar la mejor ruta para lograrlo (Wyse, Hayward y Pandya, 2016). Es así que el docente cuente con un mapa al que es importante darle el uso para el que fue creado, ya que contiene las especificaciones, la estructura detallada y la secuencia de las acciones educativas del programa. Otro aspecto que no hay que perder de vista es el alineamiento vertical del currículo, de un nivel a otro, de un curso a otro, así como su filosofía, vocabulario académico especializado, conexión interdisciplinaria y el uso de recursos y herramientas digitales. También son deseables la inter y la transdisciplina, ya que las concepciones tradicionales del currículo difícilmente se integran al privilegiar los contenidos de la disciplina particular de la carrera de que se trate, aun cuando es una de las condiciones más relevantes para el aprendizaje (Martone y Sireci, 2009; Wijngaards-de Meij y Merx, 2018).

El currículo es uno de los elementos fundamentales de cualquier sistema educativo, por lo que requiere una revisión frecuente para su mejora y en la medida de lo posible minimizar inconsistencias, así como la conjunción de resultados de aprendizaje claros y nutrirse de las necesidades de la sociedad (Ainsworth y Donovan, 2019). A las autoridades educativas y cuerpos académicos les corresponde planear y diseñar un mejor futuro para los alumnos, mediante la alineación de los distintos factores que intervienen en la educación. De esta manera el currículo, la evaluación y los métodos de enseñanza se deben colocar al centro

para la generación del cambio. Un currículo balanceado comprende evaluaciones periódicas que le permiten al alumno conocer desde el inicio del curso cómo y de qué tipo serán las evaluaciones que lo acompañan. Es importante destacar que las evaluaciones finales o sumativas también requieren de una planeación y diseño alineados con el currículo y los métodos de enseñanza. El currículo y la evaluación son componentes que se integran en cualquier ambiente de aprendizaje, sea presencial o a distancia. El currículo es el contenido y la evaluación da cuenta del progreso del aprendizaje, de lo enseñado y lo aprendido en el curso.

Las siguientes son algunas recomendaciones para el alineamiento del currículo:

1) La primera pregunta es ¿en dónde están mis alumnos?, ello con la intención de fijar el punto de partida.
2) ¿Todos los contenidos del programa necesitan ser revisados en la misma medida o hay que priorizar?
3) Considerar los recursos que se tienen, ¿cuento con los recursos necesarios –tiempo, condiciones físicas y tecnológicas–?
4) ¿Qué se podría mantener o agregar al currículo?, compartir estas inquietudes con los responsables del plan de estudios, a fin de que sirva como retroalimentación en futuras actualizaciones.
5) Identificar ideas preconcebidas sobre el tema, a fin de enfrentarlas. Por ejemplo, la creencia de que priorizar significa eliminar contenidos, la identificación de los contenidos fundamentales no exime al docente de enseñar lo que está en el currículo. Otra creencia es que priorizar es enseñar lo mínimo que debería saber o saber hacer el alumno; no se trata únicamente de eso, sino identificar qué es lo fundamental para el aprendizaje.

EVALUACIÓN

Cuando hablamos de la evaluación del aprendizaje es importante hacer una clara distinción en cuanto a los propósitos, usos y beneficios que proporciona el llevarla a cabo en los diferentes escenarios de las instituciones educativas. Sin lugar a dudas es un reto profesional para los docentes, porque involucra transformar los métodos de enseñanza y la evaluación tradicional dado que las tendencias actuales invitan a ver el aprendizaje como un proceso dinámico, creativo y flexible. La evaluación *del*, *para y como* aprendizaje es un potenciador del aprendizaje y de cambios en los métodos de enseñanza (capítulo 1).

Black y Wiliam (1998) en su artículo "*Inside the Black Box: Raising Standards Through Classroom Assessment*", constituyeron un detonante para que se realice la distinción de evaluación *del* y *para* el aprendizaje en los escenarios educativos, que no solo se hable de evaluación del aprendizaje, así como de los beneficios que se obtienen en el aprendizaje al realizar la evaluación. La evaluación a gran escala durante décadas ocupó un lugar preponderante en los escenarios educativos, cuyo máximo desarrollo implicó dejar en segundo plano la evaluación que los profesores realizan cotidianamente. De acuerdo con Shepard (2000),

la rendición de cuentas mediante la evaluación con pruebas objetivas desprofesionalizó y minimizó las habilidades de los profesores para evaluar.

La investigación en el campo de la evaluación enfatiza que la evaluación *del* y *para* el aprendizaje son elementos cruciales para apoyar a los alumnos en su proceso de aprendizaje (Earl, 2013; Saunders et al., 2011). Se ha encontrado que el efecto de la evaluación *para* el aprendizaje sobre el rendimiento de los alumnos es cuatro o cinco veces mayor que el efecto de la reducción del tamaño de la clase (Ehrenberg, Brewer, Gamoran, y Willms, 2001, como se citó en Stiggins et al., 2007) y que pocas intervenciones en educación se acercan a tener el mismo impacto (Stiggins, Arter, Chappuis, J y Chappuis, S., 2007). La evaluación puede verse como un proceso de generación y recolección de datos, los cuales se interpretan para emitir un juicio, que a su vez es comunicado y empleado para la toma de decisiones educativas.

El cambio de perspectiva sobre el papel que desempeña la evaluación en los escenarios educativos no solo consiste en considerarla como una herramienta poderosa, también pretende invitar a los docentes a cuestionar las creencias que tienen sobre sus prácticas educativas y a emplearlas de forma más integral y deliberada. Desde esta perspectiva, la evaluación es un pilar del proceso educativo. Se ha documentado que los docentes que están involucrados en implementar evaluaciones, a menudo se centran en el uso de técnicas de evaluación de manera superficial; en un reporte solo el 20% de los profesores utilizaron la evaluación en la forma en que fue diseñada para ayudar a los alumnos en el desarrollo de su aprendizaje (James et al., 2007, como se citó en Earl, 2013).

La evaluación puede tener diferentes usos, por ejemplo, los docentes la pueden emplear para que sus alumnos potencien su aprendizaje y obtengan los resultados de aprendizaje esperados, para verificar que una estrategia de enseñanza cumple el objetivo para el que se empleó, para rendición de cuentas, entre otros. La evaluación *del* aprendizaje generalmente tiene lugar al final de la instrucción y permite contar con evidencias de que los alumnos aprendieron durante el curso. En los espacios educativos es común que las evaluaciones *del* aprendizaje se realicen para asignar una calificación, aunque es importante destacar que esta no es su única finalidad (capítulo 5). Una buena evaluación *del* aprendizaje debe proporcionar información útil sobre el rendimiento del alumnado en el curso, estar vinculada con los objetivos de aprendizaje previstos, medir lo que se supone que debe medir y ser justa, esto es, brindar las mismas posibilidades de éxito a todos los alumnos (SERVE, 2006).

La evaluación *para* el aprendizaje tiene lugar durante el proceso de aprendizaje, y no al final del curso. Algunos de sus propósitos son: retroalimentar a los alumnos sobre el progreso de su aprendizaje, proporcionar al docente información pertinente para que modifique sus métodos de enseñanza, estimular la autorregulación e independencia de los alumnos en su proceso de aprendizaje, entre otros (capítulo 1). Se puede decir que la retroalimentación es esencial en la evaluación *para* el aprendizaje y se considera una de sus principales fortalezas (Christodoulou, 2016). La retroalimentación facilita al alumno verificar su progreso en términos de aprendizaje, si no la hubiera el alumno podría seguir cometiendo los mismos errores; para que la retroalimentación sea efectiva los docentes deben propiciar un clima de confianza con el grupo y establecer reglas que favorezcan la dinámica grupal. Por otra

parte, el uso de la tecnología también apuntala la diversificación de medios para evaluar con propuestas innovadoras, por ejemplo, grabación de videos, podcasts, aportaciones en wikis, en las que se privilegia la evaluación de las habilidades de comunicación de los alumnos, a diferencia de solo verificar la adquisición de conocimiento. El espectro de posibilidades con el uso de recursos tecnológicos para la evaluación se ha expandido y facilita la tarea de los docentes.

Algunas recomendaciones para el alineamiento de la evaluación:

1) Pensar en la evaluación desde el inicio de la planeación de la enseñanza, no dejarla al final del proceso. El currículo debe contar con resultados de aprendizaje explícitos, para que el diseño de la evaluación se convierta en un elemento central.
2) Compartir con los alumnos los resultados de aprendizaje esperados antes de comenzar la enseñanza, para que ellos tengan claridad respecto a dónde se espera que lleguen.
3) Planear los instrumentos y estrategias de evaluación a ser usados en el curso, con el fin de determinar cuáles serán las evidencias a generar durante el proceso y al final del mismo.
4) Incorporar una variedad de diferentes instrumentos y estrategias de evaluación del aprendizaje, que sean pertinentes para la evaluación de los resultados.
5) Analizar los resultados de la evaluación y realizar inferencias sobre el aprendizaje de los alumnos a nivel individual y grupal.
6) Usar con prudencia y asertividad la retroalimentación. Esta es más efectiva cuando es oportuna, cuando es específica sobre el trabajo del estudiante y no sobre el individuo, y cuando se encuentra alineada a criterios establecidos.
7) Emplear la información que arrojan las evaluaciones para revisar o mejorar la forma de enseñar, así como realimentar el currículo.
8) Incorporar la evaluación formativa y sumativa en congruencia con el currículo y los métodos de enseñanza. La balanza puede inclinarse a un tipo específico de evaluación, esto no quiere decir que una sea más importante que la otra, cada una tiene su lugar y función.

MÉTODOS DE ENSEÑANZA

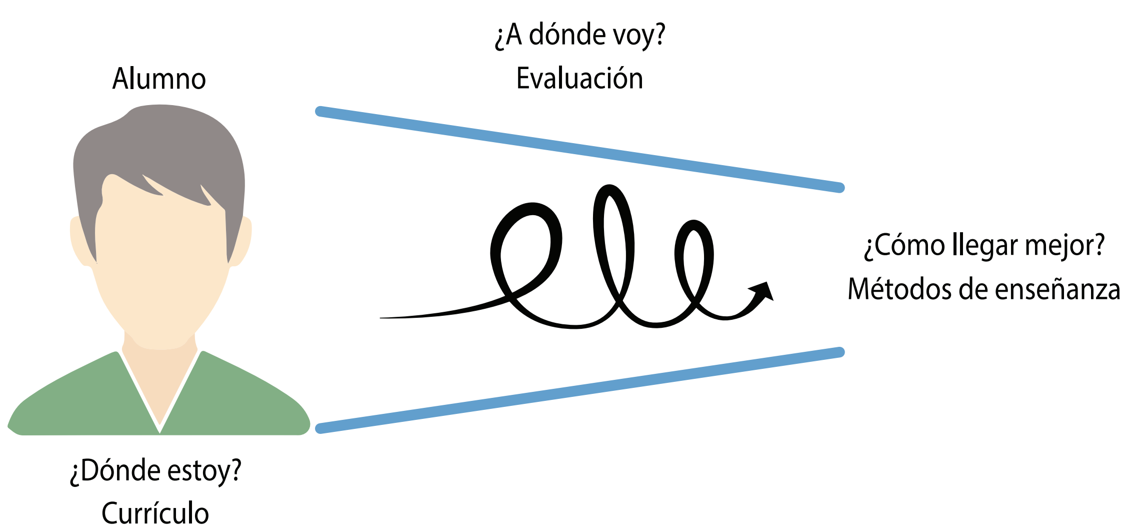
Algunos docentes se enfocan en cómo enseñar y pierden de vista el logro del aprendizaje, incluso invierten gran parte de su tiempo pensando qué van a hacer en clase, qué materiales van a emplear, qué preguntas van a hacer, en lugar de ocuparse en cómo facilitar el aprendizaje. Por ejemplo, cuando nos enfocamos en los contenidos y no en los resultados de aprendizaje, es probable que desarrollemos una buena lección, pero debemos preguntarnos cuál es su propósito, por qué pedir a los alumnos que lean tal o cual libro o que vean una película. ¿Cómo hacer que los alumnos aprendan lo que en teoría deberían aprender?, ¿qué tan realista es enseñar para aprender en el contexto tradicional del currículo con una gran cantidad de contenidos, dejando a los exámenes como recurso principal de evaluación? Aquí

radica la importancia del alineamiento con el currículo y la evaluación, ya que la interacción y balance entre ellos proveerá al alumno de un marco curricular, oportunidades balanceadas de aprendizaje, y evaluaciones *ad hoc*.

Antes de diseñar actividades para el aprendizaje, hay que pensar como facilitador del mismo. Debemos identificar los resultados de aprendizaje, así como el tipo de evidencias que se requieren para saber si los alumnos alcanzarán estos resultados; pensar en la profundidad y variedad de actividades de enseñanza que serán más apropiadas para llegar a donde se pretende (Greer, 2002; Morley y Golam, 2021; Sambell et al., 2017). Cabe destacar que la planeación de la instrucción, la selección de los métodos de enseñanza, la secuencia de las lecciones y el uso de recursos físicos y tecnológicos, tienen grandes posibilidades de ser exitosos solo después de identificar los resultados de aprendizaje, así como la determinación de los instrumentos o estrategias de evaluación que contestan las preguntas: ¿qué?, ¿a quién?, ¿para qué?, y ¿cómo? Al tener claras las metas de aprendizaje se facilita que "el aprendizaje sea visible", expresión empleada por autores como Hattie (2009) y Wijngaards-de Meij (2018).

Detrás del diseño de los métodos de enseñanza se encuentra el análisis de cómo dar sentido a la ruta para llegar al punto al que queremos llegar. Cuando hablamos del currículo enfatizamos la interrogante ¿de dónde partimos?, en el caso de la evaluación se debe enfatizar la pregunta ¿a dónde tenemos que llegar?, y en los métodos de enseñanza ¿cuál es la ruta más efectiva y más eficiente para llegar a nuestro destino? (Figura 3).

Figura 3. Perspectiva del estudiante sobre las relaciones entre currículo, evaluación y métodos de enseñanza



Debemos planear la evaluación antes de que se acerque la conclusión del curso, unidad, o tema, y dejar de confiar exclusivamente en los exámenes como el instrumento principal mediante el cual podemos verificar si los alumnos aprendieron. Existe necesidad en las instituciones educativas por mantener en primer plano el aprendizaje y considerar que la enseñanza facilita el aprendizaje, y mediante ésta se logre el aprendizaje de los alumnos. Desde esta concepción para que una intervención educativa sea considerada "exitosa" debe demostrar una mejora en el aprendizaje de los alumnos, es necesario ser más exigentes con los métodos de enseñanza que usamos. La propuesta de Hattie (2009) sobre el tamaño del efecto ha sido criticada en el ámbito educativo, ya que por sí misma cualquier intervención puede producir un efecto mayor a cero. Sin embargo, su aportación radica en evidenciar que hay intervenciones educativas que tienen mayor efecto que otras. ¿Por qué continuar con los métodos de enseñanza de siempre, si existen otros que podrían tener mayor impacto en el aprendizaje de los alumnos? Las oportunidades de aprendizaje deben ser lo suficientemente desafiantes y atractivas, lo cual contribuye a la permanencia de los alumnos. El aprendizaje "visible" (Hattie, 2009) ocurre cuando:

- El aprendizaje es la meta explícita.
- Es apropiadamente desafiante.
- Tanto el profesor como el estudiante buscan saber en qué medida se logra la meta.
- Existe una práctica deliberada dirigida a lograr el dominio de los conocimientos y habilidades.
- Hay retroalimentación efectiva y apropiada.
- Se cuenta con personas activas, apasionadas y comprometidas que participan en el acto de enseñar y aprender.

Los docentes debemos ser críticos sobre el efecto que tenemos como facilitadores del aprendizaje de los alumnos. Es fundamental hacer esta distinción, ya que con frecuencia se asume que el trabajo del docente es esencialmente práctico y se le presta poca atención al aspecto teórico, el cual es importante para entender la práctica docente. ¿Por qué o en qué se basa el docente para implementar una estrategia de enseñanza determinada? El desarrollo académico sobre el ejercicio docente se encuentra en crecimiento constante, a menudo hay incertidumbre respecto a las prácticas de investigación y de desarrollo destinadas a la profesionalización de la enseñanza. Parte del éxito del desarrollo académico es entender que los fundamentos teóricos sobre las prácticas de enseñanza facilitan la comprensión del ejercicio de las prácticas actuales y futuras. De ahí lo importante que los docentes adquieran herramientas útiles que contribuyan al fortalecimiento de sus conocimientos teóricos y habilidades prácticas, a fin de que no solo mejoren en su práctica docente, sino que influyan en el aprendizaje y éxito académico de los alumnos.

Después de más de un siglo de investigación sobre el aprendizaje en el campo de la Psicología, con trabajo intenso sobre los métodos necesarios para tener mejores resultados educativos en función de las prácticas docentes, aún se lucha en contra de prácticas

que persisten a pesar de que existe evidencia de que su impacto es mínimo, como el caso de subrayar o releer textos, entre otras (Harrington, Beale, Fancourt y Lutz, 2021). Cruzar el puente entre la teoría y la práctica debe hacerse con cuidado y recabando la mejor evidencia posible, debemos ser consumidores cuidadosos y perspicaces de la investigación educativa (Benavides Lara et al., 2021).

Una buena práctica de enseñanza procura el aprendizaje activo de los alumnos, aprender haciendo, lo cual apoya y permite a los alumnos pensar qué están haciendo y por qué pueden hacerlo. Desde esta perspectiva se requiere que los alumnos estén comprometidos e involucrados con las actividades propuestas por el docente, quien tendrá el reto de apoyar al alumno para que esto suceda. En la actualidad ofrecer información a los alumnos durante el curso es complicado debido a la gran cantidad de recursos que se encuentran disponibles en Internet, muchos de ellos gratuitos y de diferentes temáticas. Sin embargo, el docente puede apoyar al alumno en la comprensión de los materiales por medio de las propuestas de aprendizaje que diseñe para el curso.

Los docentes comprometidos con el aprendizaje activo:

- Asumen el papel de facilitadores, guías, supervisores y entrenadores.
- Hacen preguntas, estimulan la discusión y son moderadores de la misma.
- Conciben al aprendizaje como un proceso interactivo.
- Realizan evaluaciones del y para el aprendizaje.

La estimulación del pensamiento creativo en los escenarios educativos es bidireccional: del profesor a los alumnos y de los alumnos al profesor, estableciendo un ciclo dinámico. En términos generales se puede decir que el pensamiento creativo es la construcción de algo más o menos novedoso u original (Rees y Newton, 2020; Retnowati et al., 2020). Sin embargo, deben cumplirse ciertas condiciones, por lo que deben ser propuestas adecuadas y pertinentes para el propósito de aprendizaje, útiles y de valor. En estos tiempos de cambios rápidos se requiere que la enseñanza sea flexible, a fin de que pueda adaptarse a nuevas necesidades y expectativas. La enseñanza creativa prepara a los profesores para el escenario presente, futuro y de emergencia, ya que puede mejorar la flexibilidad y respuesta al cambio.

Algunas recomendaciones para la alineación de los métodos de enseñanza:

1) Los docentes deben pensar en las características de los alumnos, tener metas explícitas para maximizar el aprendizaje mediante métodos de enseñanza apropiados.
2) Enfocarse en el aprendizaje con creatividad.
3) Evitar centrarse en la mera cobertura del currículo. La principal preocupación no debe ser cumplir con el programa en el tiempo establecido, sino el aprendizaje del estudiante.
4) Determinar qué recursos y herramientas son los mejores para acompañar el logro de los aprendizajes.

5) Investigar sobre el tamaño del efecto de las estrategias de enseñanza que se proponen. Hattie (2009) en su meta análisis sobre las prácticas educativas sintetizó muchas de ellas.
6) Ofrecer experiencias de aprendizaje con tareas auténticas que representen problemas relevantes de la vida real, en las que los alumnos puedan aplicar lo aprendido.

PREGUNTAS PARA FACILITAR EL PROCESO DE ALINEAMIENTO

Para poder lograr un mejor alineamiento entre los elementos descritos del proceso educativo, es pertinente hacer las siguientes preguntas:

1) ¿Cómo estará estructurado el grupo responsable del programa educativo?
2) ¿Cuáles serán los objetivos o metas del programa?, ¿cuál será el modelo educativo utilizado?
3) ¿Qué recursos se necesitarán para el funcionamiento adecuado del programa?
4) ¿Qué barreras u obstáculos se enfrentarán?
5) ¿Qué estructura organizacional es necesaria para que se logren los objetivos?
6) ¿Son claros los resultados educativos del programa?
7) ¿Existe consenso en el profesorado sobre estos resultados?
8) ¿Existe apoyo de los líderes de la institución para el currículo y su implementación?
9) ¿Cuáles son los recursos humanos necesarios para apoyar el currículo?
10) ¿Está el programa educativo organizado para lograr los resultados descritos?
11) ¿Son compatibles los resultados de aprendizaje con la misión de la universidad y las necesidades de la sociedad?
12) ¿Cómo se evaluará el progreso de los estudiantes?
13) ¿Cuáles son los indicadores a utilizar en la evaluación del aprendizaje?, ¿internos o externos?
14) ¿Qué tipo de información de los estudiantes será identificada, recogida y analizada?
15) ¿A quiénes y cómo se les proporcionará esta información?, ¿cómo se cuidarán los aspectos éticos y de privacidad de los datos?
16) ¿Qué indicadores se monitorizarán durante el proceso, cómo se identificarán y aplicarán para realimentar al sistema?
17) ¿Cuáles pueden ser consecuencias inesperadas (positivas o negativas) del currículo en profesores, estudiantes e institución?

CONCLUSIONES

Alinear el currículo, la evaluación, y los métodos de enseñanza es una forma propositiva y cuidadosa de actuar en el ámbito educativo, la razón por la cual debe realizarse es para asegurar el aprendizaje de los alumnos y que puedan seguir su trayectoria escolar de la mejor forma. Es un proceso deliberado, planeado, reflexivo, dinámico e iterativo que tiene como brújula el aprendizaje del estudiantado, al tiempo que contempla el contexto, la cultura y los recursos locales de la institución y del profesorado.

Al realizar este alineamiento, los docentes pueden contar con evidencias para analizar, hacer ajustes a la enseñanza e implementarlos en el curso o en cursos futuros. La transformación del proceso de enseñanza–aprendizaje es un gran reto que requiere profesionalización en este campo por parte de los implicados: autoridades educativas, docentes, interesados en la evaluación y los mismos alumnos que han transitado a lo largo de experiencias tradicionales en su formación académica.

El currículo es una herramienta que, si se emplea para mejorar, puede contribuir a la institucionalización de aplicaciones pertinentes de las teorías del aprendizaje. Ello provee una estructura que se convierte en elemento trascendental para mejorar.

Implementar un sistema de evaluación integral con evaluaciones para el aprendizaje y del aprendizaje, proporciona elementos para inferir y contar con evidencias sobre el progreso de los alumnos y hacia dónde dirigirse en la experiencia educativa.

Los beneficios del alineamiento del currículo, la evaluación y los métodos de enseñanza son la coherencia en las estrategias y acciones educativas dirigidas hacia un mismo fin: el aprendizaje de los alumnos. Ello conduce a una mejora continua, ya que identifica áreas de oportunidad que permitan contribuir en la calidad educativa, así como formar alumnos más involucrados con su aprendizaje que tomen responsabilidad del mismo.

REFERENCIAS

Ainsworth, L., Donovan, K. (2019). *Rigorous Curriculum Design.* USA: International Center for Leadership in Education. Houghton Mifflin Harcourt.

Asobasi, I. (2019). *Linking Curriculum, Teaching, Assessment and the World of Work.* USA: LAPLAMBERT Academic Publishing.

Bamber, V., Trowler, P., Saunders, M., Knight, P. (2009). *Enhancing Learning, Teaching, Assessment and Curriculum in Higher Education.* Reino Unido: Open University Press. McGraw Hill Education.

Benavides Lara, M. A., de Agüero Servín, M., Pompa Mansilla, M., Sánchez Mendiola, M. (2021) El curso en Educación Basada en Evidencias (EBE): reflexiones para la transdisciplina, la docencia y la investigación. DIDAC, (78 JUL-DIC), 8-19 https://doi.org/10.48102/didac.2021..78_JUL-DIC.73

Biggs, J. B., Tang, C. (2011). Teaching for quality learning at university (4th ed.). Open University Press.

Black, P., Wiliam, D. (1998) Assessment and Classroom Learning, *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7-74, DOI: 10.1080/0969595980050102

Christodoulou, D. (2016). *Making good progress? The future of Assessment for Learning.* Reino Unido: Oxford.

Earl, L.M. (2013). *Assessment as learning*. USA: Corwin.

Greer, R. D. (2002). Designing teaching strategies: an applied behavior analysis systems approach. Amsterdam: Academic Press.

Harrington, J., Beale, J., Fancourt, A., Lutz, C. (2021). *The 'BrainCanDo' Handbook of Teaching and Learning.* London: Routledge Taylor & Francis Group.

Hattie, J. (2009). Visible learning: a synthesis of 800 meta-analyses on achievement. London: Routledge.

Martone, A., Sireci, S. G. (2009). Evaluating Alignment Between Curriculum, Assessment, and Instruction. *Review of Educational Research*, *79*(4), 1332–1361. https://doi.org/10.3102/0034654309341375

Morley, D., Golam, J. (2021). *Applied Pedagogies for Higher Education Real World Learning and Innovation across the Curriculum.* Suiza: Palgrave Macmillan.

Pendergast, D., Bahr, N. (2005). *Teaching middle years: rethinking curriculum, pedagogy and assessment.* Australia: Allen & Unwin.

Rees, S., Newton, D. (2020). *Creative Chemists: Strategies for Teaching and Learning.* Reino Unido: Royal Society of Chemistry.

Retnowati, E., Suprapto., Mohammad, A., Sugiyarto, K., Wagiron. (2020). *Innovative Teaching and Learning Methods in Educational Systems.* Reino Unido: Routledge Taylor & Francis Group.

Sambell, K., Brown, S. Graham, L. (2017). Professionalism in practice. Key Directions in Higher Education Learning, Teaching and Assessment. Reino Unido: Palgrave Macmillan.

Saunders, M., Trowler, P. Bamber, V. (2011). *Reconceptualising Evaluation in Higher Education.* USA: McGraw Hill.

SERVE Center at the University of North Carolina at Greensboro. (2006). *How to assess student performance in History: going beyond multiple-choice tests.* Recuperado de https://serve.uncg.edu/wp-content/uploads/2017/09/AssessHistory.pdf

Shepard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, *29*(7), pp.4-14. Recuperado de http://journals.sagepub.com/doi/pdf/10.3102/0013189X029007004

Stiggins, R., Arter, J., Chappuis, J., Chappuis, S. (2007). Classroom assessment for student learning. New Jersey: Pearson Education.

Webb, N. L. (1997). Criteria for alignment of expectations and assessments in mathematics and science education (Research Monograph No. 6). Washington, DC: Council of Chief State School Officers.

Wijngaards-de Meij, L., Merx, S. (2018) Improving curriculum alignment and achieving learning goals by making the curriculum visible. *International Journal for Academic Development*, *23*(3), 219-231, DOI: 10.1080/1360144X.2018.1462187

Wyse, D., Hayward, L., Pandya, J. (2016). *The SAGE Handbook of Curriculum, Pedagogy and Assessment.* Reino Unido: SAGE Publications Ltd.

Capítulo 37

LA EVALUACIÓN DE COMPETENCIAS EN EDUCACIÓN SUPERIOR: ENFOQUES TRADICIONALES Y ESTRATÉGIAS INNOVADORAS

Benilde García Cabrero, Anisai Ledesma Rodea

"La evaluación es el motor del aprendizaje, ya que de ella depende
tanto qué y cómo se enseña, como qué y cómo se aprende.
No me interesa cómo evaluar las competencias,
me interesa cómo la evaluación ayuda al desarrollo de las competencias."
Neus Sanmartí

INTRODUCCIÓN

Es ya una afirmación común entre los investigadores y profesionales de la educación (Bergsman et al., 2015), que en las últimas tres décadas ha habido un cambio gradual de enfoque, desde la educación centrada en el profesor a la educación centrada en el alumno (Reynolds y Miller, 2013). De manera concurrente, también se ha observado un cambio en el contenido de los planes de estudio, desde los centrados en los objetivos y las asignaturas, hacia los currícula centrados en competencias (Wesselink, Dekker-Groen, Biemans y Mulder, 2010).

Este cambio está relacionado con la evolución de la conceptualización del aprendizaje; la mayor parte de las teorías lo conciben como la capacidad de recordar aspectos clave de los temas impartidos, o incluso detalles de los mismos y la memorización de secuencias o procedimientos. Sin embargo, se ha pasado de esta definición a considerarlo como la capacidad de recordar, actuar y cumplir con estándares de desempeño, lo cual, representa un salto hacia la definición de un marco de referencia de aprendizaje y enseñanza basados en competencias (O'Sullivan y Bruce, 2014).

Pero ¿qué es una competencia? En diversos contextos, en particular el europeo, las competencias se consideran integradas por tres componentes: a) el conocimiento (que implica la comprensión del mismo), b) la conducta (repertorio conductual), y c) el componente valoral (que incluye valores, creencias y actitudes).

De acuerdo con el Programa Internacional para la Evaluación de los Estudiantes (PISA, por sus siglas en inglés), una competencia es más que solo conocimiento y habilidades. Implica la capacidad de cumplir demandas complejas, basándose en la movilización de recursos psicosociales (incluyendo habilidades y actitudes), en un contexto particular. Por

ejemplo, para PISA, la habilidad de comunicarse de manera efectiva constituye una competencia (comunicativa).

Con base en lo anterior, una persona competente al realizar una tarea demostrará una combinación de habilidades, conocimientos, actitudes y comportamientos, necesarios para desempeñar dicha tarea. La competencia, por tanto, puede definirse desde una perspectiva holística de estos componentes, como la capacidad de una persona para mostrar:

1) Un comportamiento particular en
2) Un contexto particular y con
3) Una característica particular.

La enseñanza basada en competencias es un tema de gran relevancia en la investigación y la práctica educativa en todo el mundo, y particularmente en la Unión Europea; prueba de ellos son los estudios PISA, PIACC y AHELO realizados por la Organización para la Cooperación y el Desarrollo Económicos (OECD, 2014). En la educación superior, esta enseñanza se ha convertido también en un objetivo de gran relevancia, en ciertos países de habla hispana, incluidos Colombia y México.

La educación basada en competencias se ha definido de múltiples formas y se ha interpretado de manera diferente en los programas académicos posteriores. Le, Wolfe y Steinberg (2014) señalan que la educación basada en competencias "es un campo en evolución sin una definición universalmente compartida de lo que hace que un modelo se base en competencias" (p. 4). Book (cit. en Gervais, 2014) argumenta que debido a que no existe una definición comúnmente aceptada, la educación basada en competencias (EBC) carece de conformidad con los estándares y de un respaldo teórico, lo que dificulta la definición clara y la implementación de manera consistente en todos los programas. A medida que se pone más énfasis en la educación basada en resultados, se ha ido creando una definición universal de EBC.

Se han utilizado diferentes términos para describir la EBC, tales como: aprendizaje basado en problemas, aprendizaje basado en el dominio, aprendizaje basado en resultados y aprendizaje basado en desempeños, pero no todos estos términos captan la esencia de la EBC, la definición de competencia operacionalizada por Gervais (2016), a partir de la revisión de la literatura y de entrevistas con informantes clave es la siguiente:

> La EBC se define como un enfoque de la educación basado en resultados, que incorpora formas particulares de impartir la instrucción y estrategias de evaluación diseñadas para evaluar el dominio del aprendizaje por parte de los estudiantes, a través de su demostración del conocimiento, las actitudes, los valores, las habilidades y los comportamientos requeridos por la profesión.

Por su parte, Pérez y Clem (2017), adoptan una definición de las competencias desde una perspectiva amplia y las conciben como:

> Desempeños multidimensionales que deben apoyarse en una base de conocimiento disciplinar y acompañarse de recursos internos (como la iniciativa y la colaboración y procedimentales (estrategias y procedimientos).

Las competencias desarrollan la habilidad de resolver problemas profesionales a medida que surgen en determinados contextos. El foco de la enseñanza de las competencias, se ubica en el "saber cómo", el desempeño se mide a través de niveles de maestría o dominio de una habilidad multidimensional. El desarrollo de las competencias se comprueba en las acciones, por lo que su enseñanza requiere de la interacción con los alumnos, más que estrategias pasivas como las que se generan durante una clase magistral o conferencia. Sin embargo, los profesores, según Pérez y Clem no han internalizado los cambios de rol que implica la EBC, por lo que existe el peligro de que este enfoque permanezca más como una retórica, que como una práctica.

Con base en lo anterior, podemos decir que la EBC representa un reto complejo, ya que integra una gran cantidad de temas de la educación contemporánea y los estudios que incluyen todos los aspectos relevantes son escasos. O'Sullivan y Bruce (2014) señalan una serie de atributos de la enseñanza basada en competencias, que incluyen aspectos importantes a considerar cuando se propone hacer realidad el enfoque de la EBC. En primer lugar, plantean que es necesario tener una comprensión clara de la forma como aprenden los alumnos, así como que *el profesor sea un modelo de humildad, pensamiento crítico, respeto, de la competencia misma que se está trabajando y de una ética del cuidado que debe demostrarse en todo momento*. Es fundamental también promover y esperar que el alumno asuma la responsabilidad sobre su propio aprendizaje, a partir de la retroalimentación constante, oportuna y específica, incluyendo la propia autoevaluación del estudiante y en la medida de lo posible, sería deseable individualizar las experiencias de aprendizaje.

Por lo tanto, una vez establecidos los perfiles profesionales de competencias en diversos programas de estudio, surgen tres preguntas relacionadas con respecto a: 1) la forma de abordar el desarrollo de las competencias (enseñanza basada en competencias: EBC) mediante diferentes métodos de instrucción, 2) qué competencias deben tener los profesores que adoptan la metodología de EBC, y 3) cómo deben evaluarse los resultados de los programas para valorar el desarrollo y la adquisición de las competencias propuestas.

CÓMO ABORDAR EL DESARROLLO DE LAS COMPETENCIAS

La manera en que los modelos de enseñanza basados en competencias se han aplicado, ha evolucionado en el tiempo. Book, (cit. en Gervais, 2016) describe dos diferentes modelos de enseñanza basados en competencias que están siendo utilizados en la actualidad. El primero es el de valoración directa, que no está ligado al material del curso y las horas-crédito, en el que los estudiantes demuestran el dominio de las competencias a su propio ritmo, generalmente a través de cursos virtuales. Los grados se otorgan con base en las demostraciones que hagan los estudiantes a través de una amplia variedad de métodos de evaluación. No

se otorgan puntuaciones, más bien se realizan evaluaciones formales para determinar si los alumnos cuentan con las competencias que se requieren para otorgar el grado. El segundo modelo, que está basado en créditos por hora, en el que las competencias se establecen a nivel del programa dentro de tópicos, que a su vez se integran en cursos de una duración y complejidad adecuadas. El aprendizaje de los alumnos en este segundo tipo de cursos, se mide por el tiempo dedicado a la tarea, el número de cursos tomados y las tareas que han recibido calificaciones. Como ocurre en los modelos tradicionales de enseñanza, las calificaciones otorgadas en los programas de créditos por hora, están basadas en el número de créditos acumulados por los alumnos y las calificaciones se basan en una escala que va de la A a la F, y eso constituye la evaluación formal del logro de un estudiante en un programa por competencias.

La adquisición y desarrollo de las competencias se apoya en proporcionar feedback y en las contribuciones de todos los involucrados, incluyendo alumnos, profesores y socios comunitarios. La complejidad de la adopción de un modelo de EBC puede corroborarse en la Tabla 1, en la que se describe el proceso de adopción de un enfoque de enseñanza basado en competencias, tomando como punto de partida un enfoque tradicional de enseñanza, su transición hacia un modelo emergente de enseñanza basada en competencias y la consolidación del mismo.

La tabla presenta los cambios que deberán irse dando en cada uno de los siguientes componentes, para hacer realidad la implantación de un modelo de EBC:

a) La cultura escolar en la institución
b) La progresión de los aprendizajes
c) El ritmo de cada estudiante para adquirir los aprendizajes
d) Las estrategias instruccionales
e) El sistema de evaluación
f) La forma de acreditar los aprendizajes

Tabla 1.El continuum de la educación basada en competencias

	Enfoque Tradicional	Enfoque Emergente	EBC
Cultura escolar	El aprendizaje ocurre dentro de un aula tradicional con poca o ninguna adaptación a los intereses y estilos de aprendizaje de los estudiantes.	Los profesores y su enseñanza se adaptan de manera limitada a los intereses y estilos de aprendizaje de los estudiantes, al incorporar experiencias y socios del mundo real en el aula.	Los estudiantes eligen entre una amplia gama de experiencias de aprendizaje en el entorno presencial, en línea y en su comunidad. Los profesores trabajan con diversos socios y estudiantes para promover cambios de aprendizaje a nivel individual que se adapten a los intereses y estilos de aprendizaje de los estudiantes.

Progresión del aprendizaje	Se espera que los estudiantes dominen los estándares de desempeño universitario y profesional a nivel de grado.	Se espera que los estudiantes dominen los estándares de desempeño profesional universitario a nivel de grado y habilidades transferibles.	Se espera que los estudiantes dominen las competencias alineadas con los estándares de preparación universitaria y profesional. Cada competencia tiene objetivos claros de aprendizaje transferibles.
Ritmos de aprendizaje	Los estudiantes avanzan en la instrucción al ritmo del tutor, independientemente de si han dominado los objetivos de aprendizaje o si necesitan tiempo adicional.	Los estudiantes pueden tomar cursos de forma acelerada si demuestran estar preparados. Los estudiantes reciben apoyo especializado cuando se quedan atrás de sus pares. Los educadores continuamente forman grupos de estudiantes para fomentar el aprendizaje entre pares y maximizar los beneficios de aprendizaje para todos.	Los estudiantes reciben apoyos personalizados y oportunidades aceleradas tanto en la escuela como fuera de la escuela, para garantizar que se mantengan al día y graduarse de la universidad y la carrera profesional.
Instrucción	Cada clase tiene un maestro que diseña y entrega un programa de instrucción con muy poca diferenciación para estudiantes individuales.	Los profesores participan en alguna forma de colaboración entre equipos y áreas de contenido para alinear y diferenciar la instrucción en función de proporcionar retroalimentación en tiempo real sobre el desempeño de los estudiantes.	Los profesores trabajan en colaboración con otros socios de la comunidad y los estudiantes para desarrollar un plan de aprendizaje para cada estudiante basado en estilos de aprendizaje y datos obtenidos en tiempo real.
Sistema de evaluación	Los instrumentos de evaluación se utilizan en momentos establecidos para evaluar y clasificar a los estudiantes, no para guiar la instrucción. Los estudiantes tienen una sola oportunidad de realizar la evaluación sumativa al final del año.	Los profesores usan instrumentos de evaluación formativa cuando creen que los estudiantes están listos para demostrar su dominio. Estas evaluaciones ayudan a los educadores a adaptar la instrucción para que más estudiantes estén listos para aprobar la evaluación sumativa al final del año.	Un sistema de evaluación integral es una parte esencial del sistema de aprendizaje. Las evaluaciones formativas guían la instrucción diaria y la selección de oportunidades de aprendizaje personalizadas por parte de los estudiantes. Las evaluaciones sumativas demuestran el dominio de las competencias. Los estudiantes toman estas evaluaciones cuando están listos y tienen múltiples oportunidades para demostrar su dominio.
Políticas de calificación	Las calificaciones se basan en normas que reflejan el dominio de los estándares del curso y, por lo general, se basan en trimestres ponderados y un examen final.	Las calificaciones reflejan el dominio de los estándares y habilidades del curso y generalmente se basan en trimestres ponderados y un examen o proyecto final. Los estudiantes tienen múltiples oportunidades para demostrar el dominio de los cursos requeridos.	Las calificaciones reflejan el grado de dominio de las competencias, que van desde avanzado hasta aún no competente. Cuando los estudiantes no obtienen créditos por el curso, su registro indica competencias que deben volver a aprender en lugar de tener que repetir todo el curso.

Tomada de Gervais (2016).

Un plan de estudios basado en competencias se centra menos en lo que los alumnos necesitan saber y más en cómo los alumnos aplican sus conocimientos, habilidades y actitudes en el mundo real. Como resultado, un plan de estudios basado en competencias ayuda a los estudiantes a desarrollar las competencias genéricas y específicas necesarias para progresar en su educación o progresar hacia la obtención de un empleo.

El rol de los alumnos en la EBC se caracteriza por:

- Construir sobre el dominio que han ido demostrando sobre el contenido teórico-conceptual y metodológico de la materia.
- Demostrar su capacidad para transferir el aprendizaje entre diferentes entornos (esto puede incluir tareas en las que se pide a los alumnos que demuestren sus conocimientos, pero no reciben una calificación final o formal, lo que se conoce como 'evaluación formativa' o evaluación para el aprendizaje)
- Evaluaciones completas que sean significativas y relevantes para las habilidades requeridas en los entornos profesionales.
- Desarrollar y aplicar las habilidades y disposiciones necesarias para un desempeño profesional exitoso en los entornos reales de trabajo (Open University, 2020).

Las competencias han sido clasificadas de diferentes maneras por diversos autores e instituciones en función de su lógica, teoría y propósito de estudio, Saquib, Malik y Farruk (2012), afirman que las formas de clasificación en la literatura más comunes son las siguientes:

- **Competencias blandas y duras.** Competencias como analizar y organizar se consideran competencias duras, es decir, competencias difíciles, mientras que la creatividad, habilidades interpersonales y de comportamiento son competencias blandas. Sin embargo, este concepto de clasificación es a menudo criticado sobre la base de que la diferenciación entre competencias duras y blandas es siempre difícil de comprender y establecer un significado conceptual.
- **Competencias de desempeño y umbral.** Las 'competencias umbral' son un requisito mínimo básico, mientras que las 'competencias de desempeño' son habilidades y competencias que realmente diferencian entre promedio y excelente desempeño. Esta distinción ha sido criticada debido a que se ha argumentados que es una cuestión de grado en lugar de una categoría.
- **Clasificación jerárquica sabia.** Se basa en la categorización de las competencias que se necesitan en tres niveles jerárquicos de gestión.
- **Clasificación previa y empírica.** La clasificación empírica presenta de alguna manera necesidades de competencias variadas a partir de la clasificación de competencias derivada teóricamente (previa).

Por su parte, Roegiers (2016), señala que las competencias entendidas como una potencialidad del individuo para actuar en una situación de manera adecuada, se pueden clasificar en tres categorías:

- **Competencias de conocimiento:** esta categoría involucra conocimientos cognitivos, manuales y técnicos, se puede representar de la siguiente manera: competencia = potencial para actuar + contenido educativo / técnico (en formación técnica o profesional), como reconocer un triángulo, comparar dos modos de germinación, montar un marco de puerta, etcétera.
- **Competencias genéricas:** estas competencias son significativas solo si están conectadas con contextos, algunos ejemplos son la mentalidad abierta, el asertividad y la creatividad. Por ejemplo, alguien puede ser muy abierto en el lugar de trabajo, pero tienen dificultades para comunicarse en la familia con los hijos adolescentes. Esta categoría se puede representar de la siguiente manera: **competencia = potencial para actuar + contexto.**
- **Competencias situacionales:** Es un conjunto de contenidos adquiridos por el alumno y que debe utilizar en diversas situaciones. Estas situaciones ("situaciones de integración") son lo suficientemente complicadas, pero no demasiado y son suficientemente complejas, pero no demasiado. Se pueden representar de la siguiente manera: competencia = potencial para actuar + contenido educativo / técnico + situación.

En cuanto a la evaluación de las competencias, según la clasificación propuesta por Roegiers (2016), se realiza de la siguiente manera:

- **Evaluación de conocimientos.** Desde la introducción de la enseñanza por objetivos en la educación, se evalúan los tipos de conocimientos en la escuela de acuerdo con procedimientos sistemáticos relativamente satisfactorios, ya sea a través de cuestionarios, ejercicios, preguntas frecuentes o práctica.
- **Evaluación de competencias genéricas.** La evaluación de competencias genéricas en la docencia cuenta con muy pocos instrumentos o dispositivos, por lo que aún son evaluados a menudo a través de una apreciación subjetiva dada por el profesor.
- **Evaluación de competencias situacionales.** Aparte de la formación técnica y profesional, donde estas evaluaciones son una práctica habitual, no existe, en el mundo de la educación general primaria o secundaria, una tradición arraigada de evaluar las competencias situacionales. Sin embargo, ya han comenzado a ser evaluadas a través de situaciones complejas presentadas al alumno: situaciones en las que se produce una presentación escrita compleja, se resuelve un problema, etc. Además, se hace uso de pruebas estandarizadas.

LAS COMPETENCIAS DE LOS PROFESORES QUE ADOPTAN LA METODOLOGÍA DE EBC

En congruencia con la adopción de un modelo de EBC, es necesario considerar la definición de los roles y funciones del docente en términos de competencias, por lo que cabe hacer la pregunta: ¿Qué son las competencias docentes? Las competencias docentes pueden definirse como las habilidades y los conocimientos que permiten a un maestro tener éxito en lograr que sus alumnos aprendan y se sientan satisfechos con la instrucción recibida, lo cual se corrobora generalmente a través de los cuestionarios de opinión de los alumnos al terminar un semestre lectivo. Para maximizar el aprendizaje de los estudiantes, los maestros deben tener experiencia en una amplia gama de competencias en un entorno de enseñanza especialmente complejo (que se ha visto agravado con la pandemia por COVID-19) donde se requieren cientos de decisiones críticas cada día (Jackson, 1990). Pocos trabajos exigen la integración del juicio profesional y el uso eficiente de competencias basadas en la evidencia como lo hace la docencia.

¿Por qué es esto importante?, es decir, ¿por qué necesitamos precisar cuáles son las competencias docentes de un profesor eficaz? El poder transformador de un maestro eficaz es algo que muchos de nosotros hemos experimentado y por supuesto los alumnos lo han reportado y existen muchos testimonios que dan cuenta de ello. Sin embargo, no es suficiente con que los alumnos reporten que los profesores han cambiado su actitud hacia una asignatura, las matemáticas, por ejemplo, sino que es necesario confirmar estas relaciones. Intuitivamente, el vínculo entre la enseñanza y el rendimiento académico de los estudiantes puede parecer obvio, pero ¿cuál es la evidencia de ello?

Diversas investigaciones han confirmado esta percepción común del vínculo y revelan que, de todos los factores bajo el control de una escuela, los maestros son la influencia más poderosa en el éxito de los estudiantes (Babu y Mendro, 2003). ¿Qué separa a los profesores eficaces de los ineficaces y cómo se puede utilizar esta información para apoyar una mejor enseñanza? Actualmente, es posible empezar a construir un perfil ejemplar de instrucción en el aula derivado de la investigación sobre la docencia eficaz (Wenglinsky, 2002; Hattie, 2008; Hattie et al., 2020) y en el caso particular de este capítulo, consideramos que es necesario desarrollar un perfil instruccional de un profesor que enseñe de acuerdo a los principios e implicaciones de la EBC.

Para que se considere que un profesor está enseñando por competencias, ya no es aceptable que solamente se presente a la clase e imparta conferencias a partir de su conocimiento sobre un tema. La EBC requiere una gran cantidad de preparación previa para impartir clases (Elbow, 1979, cit. en Gervais, 2016). La planificación requiere pensar en el propósito de una clase, tomando en cuenta el contexto de todos los cursos, diseñar y llevar a cabo una instrucción alineada con los objetivos de aprendizaje y decidir cómo impartir el programa de estudios del curso, considerando los aspectos teóricos y los prácticos, así como la modalidad, es decir, en línea, presencial o de forma híbrida. En un modelo de enseñanza de EBC, el aprendizaje se estructura de forma horizontal y vertical (Albanese et al., 2008). El aprendizaje horizontal significa que los estudiantes deben aprender a integrar lo que aprenden a lo largo de todo el plan de estudios. El aprendizaje vertical significa que el estudiante debe

dominar el contenido de cada curso en profundidad. Un estudiante en un modelo EBC debe demostrar el dominio tanto de los contenidos considerados de forma vertical dentro de su asignatura y en relación con las otras asignaturas de un semestre, módulo o bloque, así como el horizontal, de un semestre a otro, con miras al logro del perfil profesional. Esto implica que los cursos no pueden enseñarse de forma separada, sino que un curso determinado debe apoyarse y extender las competencias logradas en el anterior y favorecer el siguiente.

La evaluación en un curso de EBC se considera un componente clave para apoyar el aprendizaje de los estudiantes y representa una oportunidad para calibrar y demostrar que han adquirido la(s) competencia(s), lo que depende en buena medida de la habilidad de los maestros para determinar tanto las necesidades de aprendizaje de los alumnos, como si los instrumentos utilizados están midiendo lo que pretenden medir, considerando que las valoraciones del desempeño no se toman como una medida para comparar a los estudiantes entre sí y las puntuaciones no se ubican dentro de una curva normal.

EVALUACIÓN DE RESULTADOS: ESTRATEGIAS TRADICIONALES E INNOVADORAS EN LA EVALUACIÓN DE COMPETENCIAS.

Las nuevas evaluaciones o la Evaluación Basada en Competencias (CBA, por sus siglas en inglés), tienen características diferentes a las pruebas escritas, estandarizadas y tradicionales que tienen como objetivo probar una base de conocimientos, por lo que se han propuesto muchas nociones teóricas para caracterizar la CBA, como centrarse en el desempeño en situaciones auténticas, combinar múltiples métodos, involucrar a diversos evaluadores preferiblemente con diferentes antecedentes, utilizar la puntuación de referencias de criterios e integrar el aprendizaje con las actividades de evaluación (Gulikers, et al., 2009).

Desafortunadamente, todavía hay poca evidencia empírica sobre la calidad de los CBA que incorporan estas características teóricas (Segers y Dochy, 2006). La investigación existente se centra en examinar características específicas como la autenticidad (Gulikers et al., 2009) o la participación de los estudiantes (Sluijsmans y Prins, 2006) en lugar de examinar un CBA en toda su amplitud y coherencia, o centrarse en los efectos de una determinada evaluación en los estudiantes (por ejemplo, Harlen, 2005). Todavía hay poca evidencia empírica que muestre qué características teóricas de CBA realmente impactan la calidad de los CBA en la práctica. Esta cuestión se complica aún más por el reconocimiento de que estas nuevas evaluaciones requieren una nueva forma de examinar su calidad.

Una estrategia tradicional de evaluación que ha sido utilizada para garantizar la "igualdad de condiciones" a los alumnos, es la estandarización de la evaluación de los aprendizajes. Esto ha permitido que todos los alumnos puedan ser evaluados en función de determinados resultados de aprendizaje necesarios para aprobar una asignatura o avanzar hacia un nivel educativo superior (Open University, 2020). Esto, aunque tiene la ventaja de que todos sean igualmente evaluados, no garantiza que todos hayan sido enseñados de la misma forma, por lo que, en ocasiones, la forma en que se evalúa a los estudiantes está desconectada

de la manera en que se les ha enseñado. Además, la evaluación no refleja fielmente lo que se les pedirá que hagan si continúan su tránsito hacia la educación superior, o se insertan en un contexto laboral específico.

Contrario a lo anterior, la evaluación por competencias, es un "proceso de recogida constante de información sobre las capacidades desarrolladas por el estudiante, de comparación de su desempeño con el perfil de competencias propuesto por un Plan de Estudio y de formulación de un juicio de valor sobre el grado de ajuste al mismo." (Sanz de Acedo Lizarraga, 2010, p. 24).

Por ejemplo, si un alumno adquiere la información que necesita para lograr los resultados de aprendizaje del curso a través de la lectura de un libro, ¿sería apropiado utilizar una presentación oral para evaluar su comprensión del libro y su contenido? Lo anterior puede ser apropiado, sin embargo, es importante considerar que se pueden realizar diferentes evaluaciones, por ejemplo:

- **Evaluación formativa:** Se le pide al alumno que exprese su comprensión del libro o un capítulo del libro como parte de una presentación a una clase.
- **Evaluación sumativa:** Después de lo anterior, se evaluará la comprensión del estudiante en un examen oral o escrito.

Al realizar estas dos formas de evaluación, el vínculo entre cómo se evaluará al alumno y cómo aprendió y demostró el aprendizaje está más alineado (Open University, 2020).

¿QUÉ ES LA EVALUACIÓN INNOVADORA Y POR QUÉ ES ÚTIL?

El uso de la evaluación innovadora se define como algo que permite al alumno transferir conocimientos, habilidades, competencias y disposiciones entre el aprendizaje y la evaluación. La transferencia entre aprendizaje y evaluación debe ser, cuando sea posible, fluida. La premisa detrás de este enfoque es garantizar que el alumno pueda ver una conexión clara entre lo que está aprendiendo y cómo se aplicará ese aprendizaje en cualquiera o en todas las siguientes áreas:

- El mundo real
- Un ambiente de trabajo
- Avance hacia un nivel superior de educación (Open University, 2020).

Para poder implementar y evaluar la utilidad de cualquier evaluación innovadora, primero debe considerar cómo la evaluación debe alinearse con el plan de estudios que está enseñando. Además, se debe considerar cómo la evaluación puede contribuir al logro de los resultados del aprendizaje en el plan de estudios de la asignatura que está enseñando. Cada plan de estudios y materia es diferente con diferentes resultados de aprendizaje y habilidades que los alumnos deben demostrar. No se trata únicamente que el alumno demuestre

conocimiento y comprensión, sino que también es importante evaluar otras áreas, incluidas las habilidades cognitivas, como la integración de evidencia para respaldar una respuesta; o habilidades prácticas, como demostrar conciencia de la salud y la seguridad en un entorno en el que puede estar trabajando. También se les puede solicitar que demuestren habilidades clave, como su capacidad para interpretar información numérica o utilizar la tecnología para acceder a las fuentes de información (Open University, 2020).

Ion, Cano y Cabrera (2016), señalan que la evaluación de competencias también debe recolectar evidencia sobre la progresión del conocimiento y brindar a los estudiantes información sobre el proceso. Es así que, la evaluación formativa se vuelve una estrategia clave para mejorar la calidad de la enseñanza y el aprendizaje. No solo se trata de proporcionar evidencia de conocimiento, sino de promover la conciencia del propio aprendizaje de los estudiantes.

En este sentido, Bierer, Dannefer, Taylor, Phillip y Hull (2009), proponen dos métodos para evaluar el conocimiento a partir de una investigación llevada a cabo en un plan de estudios basado en competencias: las preguntas tipo ensayo y preguntas de opción múltiple, dichos métodos se utilizaron para proporcionar a los estudiantes comentarios formativos semanales sobre sus habilidades para adquirir, aplicar e integrar conceptos básicos y esenciales para la práctica.

Preguntas tipo ensayo

Las preguntas tipo ensayo estaban diseñadas para desafiar la capacidad de un estudiante para integrar y aplicar los conceptos básicos presentados con anterioridad. Los estudiantes recibían de dos a tres preguntas tipo ensayo los días miércoles y debían enviar sus respuestas electrónicamente el lunes siguiente. Estas preguntas tipo ensayo consistían en presentar un caso clínico y posteriormente realizar diferentes cuestionamientos sobre el caso planteado. Una vez que los alumnos enviaban sus respuestas, se les brindaban dos retroalimentaciones para cada contestación: una ‘respuesta’ estándar y una retroalimentación individualizada. Las respuestas estándar se publicaban en el portal de la escuela después de la fecha límite de envío para ayudar a los estudiantes a evaluar la precisión y exhaustividad de sus respuestas y enfocar el estudio independiente de conceptos poco claros durante la semana siguiente. Adicionalmente, un miembro de la facultad leía todas las respuestas y proporciona a cada estudiante comentarios narrativos individualizados dentro de los siete días posteriores a la fecha límite de entrega. Esta retroalimentación se agregaba automáticamente al portafolio electrónico de cada estudiante, que también incluía la respuesta del estudiante a la pregunta para proporcionar documentación de la “profundidad” del conocimiento sobre el tema.

Preguntas de opción múltiple

Esta segunda medida del conocimiento utilizó elementos de opción múltiple diseñados por los profesores del curso para ayudar a los estudiantes a monitorear su comprensión de conceptos importantes en el plan de estudios. Por tanto, las preguntas de autoevaluación funcio-

naban como medidas formativas de la “amplitud” del conocimiento. Los estudiantes completaban aproximadamente 30 preguntas de autoevaluación cada semana para identificar su adquisición de conceptos clave abordados esa semana en el plan de estudios. Debido a que un entorno de prueba seguro es innecesario para estas evaluaciones formativas, los estudiantes reciben puntajes brutos y una retroalimentación inmediatamente después de completar las preguntas de forma electrónica. Esta retroalimentación oportuna tiene como objetivo ayudar a los estudiantes a determinar si captaron los principales conceptos enfatizados esa semana. Además, los directores de los cursos recibían informes semanales del desempeño total de los estudiantes mismos que los ayudaban a identificar conceptos que la clase no captó o detectar elementos que requieren revisión.

Los autores afirman que los dos métodos anteriores, proporcionan una manera relativamente fácil de otorgar a los estudiantes una responsabilidad adicional para monitorear su aprendizaje más allá de la dependencia de libros de texto o bancos de artículos comerciales destinados a la preparación de exámenes de licenciatura. El acto de crear preguntas de opción múltiple también ayuda a los profesores a identificar y priorizar los conceptos básicos para que los estudiantes aprendan a lo largo de un curso, siempre y cuando las preguntas estén bien redactadas y alineadas con los objetivos de aprendizaje.

La evaluación de competencias haciendo uso de las tecnologías digitales ha sido abordada por organismos como la UNESCO (https://es.unesco.org/news/evaluacion-competencias-contexto-pandemia-webinar) que señala que si se atiende al principio básico que define el aprendizaje basado en competencias, en los entornos virtuales, es posible promover en los estudiantes el desarrollo de la transferencia de estos aprendizajes. En este sentido, una propuesta para evaluar las competencias de los estudiantes universitarios haciendo uso de herramientas digitales es la realizada por Ion, Cano y Cabrera (2016), quienes crearon una plataforma que permitió a los profesores registrar sus comentarios y evaluaciones a través de blogs creados por los estudiantes. La plataforma fue nombrada Herramienta de Evaluación de Competencias (CAT, por sus siglas en inglés), dicha plataforma permitía a cada profesor, identificar las competencias a evaluar y los indicadores que la conformaban. Posteriormente, los docentes decidían qué actividades específicas debían desarrollar los alumnos, mismas que se llevaban a cabo en formato blog (libre y de acceso abierto). Posteriormente, los estudiantes se unían a la plataforma y creaban sus blogs, etiquetando cada entrada según los indicadores asignados. Finalmente, cada docente brindaba retroalimentación periódica sobre el trabajo realizado utilizando la plataforma CAT. Un sistema de lista de verificación aseguró la relevancia y validez del proceso de evaluación, lo que les permitía acceder a los indicadores de competencia que los estudiantes habían publicado y evaluar si se habían logrado (azul) o no (rojo), y proporcionar retroalimentación a los estudiantes sobre el nivel de las competencias adquiridas.

Una de las principales ventajas que se identificaron sobre el uso de la plataforma CAT, es su capacidad para ofrecer continuamente a los estudiantes y profesores una imagen global de las competencias que necesitan alcanzar. Además, permitió a los estudiantes tomar conciencia de lo que implica el proceso de aprendizaje y aumentar su autonomía.

La Open University (2020), señala las habilidades y competencias que un alumno podría demostrar junto con ideas de evaluación innovadoras que podrían usarse para desarrollar estas habilidades y también ayudar al estudiante a desarrollar conocimientos. En la Figura 1, se pueden observar diversas maneras de promover la orientación y el apoyo hacia los estudiantes, por ejemplo, haciendo uso del andamiaje, propiciando el apoyo entre los estudiantes del grupo, el aprendizaje por descubrimiento, entre otros. Asimismo, se pueden visualizar algunas características que es importante integrar respecto de los contenidos revisados y la experiencia que tengan los estudiantes con éstos, como, por ejemplo, que sea de fácil acceso, basado en la investigación y en la práctica, así como que brinde ejemplos a los estudiantes sobre la aplicación de estos conceptos en situaciones cercanas a su contexto. En cuanto a la evaluación y al trabajo colaborativo, en la Figura 1 se muestra que es importante promover la autoevaluación, así como la evaluación entre los miembros del grupo, documentar el progreso de los alumnos mediante una carpeta de trabajo, entre otras. Adicionalmente, para promover la comunicación y el trabajo colaborativo, es importante promover el debate, así como que los estudiantes formen una comunidad de aprendizaje en la que generen conocimiento de manera colaborativa.

Figura 1. Habilidades y competencias propuestas por la Open University (2020) y su evaluación



Adaptada de Open University (2020).

CONCLUSIÓN Y REFLEXIONES FINALES

Los programas de educación basados en competencias pueden representar un camino alternativo en el que los estudiantes progresen a su propio ritmo y demuestren el dominio de competencias importantes (McClarty y Gaertner, 2015). Debido a lo anterior, es importante que al evaluar las competencias de los estudiantes también se considere realizar investigaciones longitudinales que vinculen estas evaluaciones con resultados relevantes de los estudiantes a largo plazo, lo que sería fundamental para establecer la validez de las evaluaciones realizadas.

La pandemia ha acelerado claramente la transición a la enseñanza y el aprendizaje virtuales, pero la EBC requiere que los alumnos aprendan a desempeñarse en contextos lo más parecidos a los contextos reales, a través de lo que se ha denominado enseñanza y evaluación auténticas, lo cual resulta difícil de lograr en los entornos virtuales. En particular, se ha señalado que la evaluación debe incorporar los recursos tecnológicos para volverse más automatizada, aliviando la carga de trabajo de los maestros, posibilitar que sea continua, permitir retroalimentar el desarrollo de las competencias de los estudiantes y evitar el plagio.

REFERENCIAS

Albanese,M.A., Mejicano, Anderson, W.M., Gruppen, L. (2008). Building a competency-based curriculum: the agony and the ecstasy. *Adv in Health Sci Educ 15*,439–454 http://doi.org/10.1007/s10459-008-9118-2

Babu, S. Mendro, R. (2003) *Teacher Accountability: HLM-Based Teacher Effectiveness Indices in the Investigation of Teacher Effects on Student Achievement in a State Assessment Program*, AERA Annual Meeting

Bergsmann, E., Schultes, M.T., Winter, P., Schober, B., Spiel, C. (2015). Evaluation of competence-based teaching in higher education: From theory to practice. *Evaluation and Program Planning 52*,1–9.

Bierer, S., Dannefer, E., Taylor, C., Phillip, H y Hull, A. (2009). Methods to assess students' acquisition, application and integration of basic science knowledge in an innovative competency-based curriculum. *Medical Teacher, 30*(7), 171-177. DOI: 10.1080/01421590802139740

Gervais, J. (2016). The operational definition of competency-based education. *The Journal of Competency-Based Education, 1,* 98–106.

Gulikers, J., Biemans, H y Mulder, M (2009). Developer, teacher, student and employer evaluations of competence-based assessment quality. *Studies in Educational Evaluation, 35,* 110-119.

Harlen, W. (2005). Teachers' summative practices and assessment for learning—Tensions and synergies. *The Curriculum Journal, 16*(2), 207–223.

Hattie, J. (2008). Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement. London: Routledge.

Hattie, J., Bustamante, V.; Almarode; J.T, Fisher, D., Frey, N. (2021), *Great Teaching by Design. From Intention to Implementation in the Visible Learning Classroom*.Corwin.

Ion, G., Cano, E y Cabrera, N. (2016). Competency Assessment Tool (CAT). The Evaluation of an

Innovative Competency-Based Assessment Experience in Higher Education. *Technology, Pedagogy and Education, 25*(5), 631- 648.

Jackson, P. W. (1990). Life in classrooms. New York, NY: Teachers College Press.

Le, C., Wolfe, R. E., Steinberg, A. (2014). The Past and the Promise: Today's Competency Education Movement. Students at the Center: Competency Education Research Series. Jobs for the Future.

McClarty, K., Gaertner, N (2015). Measuring mastery best practices for assessment in competency-based education. *AEI Series on Competency-Based Higher Education,* 1 – 21. https://files.eric.ed.gov/fulltext/ED557614.pdf

Organisation for Economic Co-operation and Development (2014). *Education.* Recuperado de: http://www.oecd.org/education/

O'Sullivan, N y Bruce, A. (2014). Teaching and learning in competency-based education. Fifth International Conference on e-Learning, eLearning 2014. At: University of Belgrade, Serbia. www.researchgate.net/publication/269810124_Teaching_and_Learning_in_Competency_Based_Education

Open University (2020). Open Learn. Exploring innovative assessment methods. https://www.open.edu/openlearn/education-development/exploring-innovative-assessment-methods/content-section-overview?active-tab=content-tab

Pérez, C.G. y Clem, S. (2017). Teaching practices at a Chilean university 3 years after conversion to competency-based education. *Competency-based Education, 2*(2).DOI: 10.1002/cbe.1054.

Reynolds, W. M., Miller, G. E. (2013). Educational psychology: Contemporary perspectives. En I. B. Weiner, W. M. Reynolds, G. E. Miller (Eds.), Handbook of psychology, educational psychology (Vol. 7, pp. 1–22). Hoboken, NJ: Wiley.

Roegiers, X (2016). A Conceptual Framework for Competencies Assessment. Current and Critical Issues in the Curriculum and Learning. UNESCO. https://neqmap.bangkok.unesco.org/wp-content/uploads/2019/08/245195eng.pdf

Saquib, Y., Malik, A y Farruk, N. (2012) The competence classification framework a classification model for employee development. *Interdisciplinary journal of contemporary research in business, 4*(1), 396-404. https://journal-archieves18.webs.com/396-404.pdf

Sanz de Acedo Lizarraga, M. (2010). *Competencias cognitivas en Educación Superior*. Narcea ediciones. Madrid, España.

Segers, M., Dochy, F. (2006). Enhancing student learning through assessment: Alignment between levels of assessment and different effects on learning. *Studies In Educational Evaluation, 32*(3), 171–179.

Sluijsmans, D., Prins, F. (2006). A conceptual framework for integrating peer assessment in teacher education. *Studies in Educational Evaluation, 32*, 6–22.

Wenglinsky, H (2002). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives 10*(12).

Wesselink, R., Dekker-Groen, A. M., Biemans, H., Mulder, M. (2010). Using an instrument to analyze competence-based study programmes: Experiences of teachers in Dutch vocational education and training. *Journal of Curriculum Studies, 42*(6), 813–829. http://dx.doi.org/10.1080/00220271003759249

Capítulo 38

EVALUACIÓN EN LOS ESPACIOS DE TRABAJO

Carlos Gutiérrez-Cirlos, Melchor Sánchez-Mendiola

"El trabajo sin amor es esclavitud."
Madre Teresa de Calcuta

INTRODUCCIÓN

La evaluación en los espacios de trabajo puede ser de gran importancia para la toma adecuada de decisiones en tres dimensiones: el público, la persona que busca el servicio y la persona que lo proporciona. Varias profesiones, en las que destaca la atención a la salud, han incorporado la evaluación en los espacios de trabajo, como una manera de mejorar el proceso educativo, el desarrollo profesional de los participantes, y la prestación de servicios. En este capítulo se hará referencia a ejemplos relacionados con la atención de la salud, de forma no limitativa, dado que este tipo de evaluación puede ser aplicada en prácticamente cualquier actividad en la que hay estudiantes y docentes en espacios laborales y de práctica profesional.

Históricamente, los desenlaces en salud se basan en la morbilidad y la mortalidad, pero en la perspectiva sistémica moderna los desenlaces para la evaluación se han modificado y ampliado: aspectos como la satisfacción del paciente, el estado funcional antes y después de un tratamiento, análisis de costo-efectividad y la evaluación de desenlaces durante el tratamiento, cada vez son más importantes. Por ejemplo, en una persona con diabetes mellitus, una evaluación del sistema de salud podría implicar en una perspectiva limitada, analizar exclusivamente la mortalidad y morbilidad de la enfermedad. Actualmente se incorporan al proceso de atención otros factores como calidad de vida, aspectos financieros, desarrollo de complicaciones relacionadas con la enfermedad y sus efectos en la familia. Lo anterior contribuye a dar tranquilidad y seguridad a la población, al sistema de salud y al médico, al documentar un desempeño adecuado en la práctica. Si un paciente conoce los indicadores mencionados antes, puede decidir, con mejor información, a qué sistema de salud acudir y con qué médico solicitar una consulta. Para el profesional de la salud, este tipo de evaluación le da tranquilidad de que fue realizada con base en desenlaces reales de su práctica individual y cotidiana.

La pandemia por COVID-19 puede tomarse como un ejemplo de la forma en la que la sociedad, ya sea público, prensa o gobierno, toma decisiones en cuanto al desempeño de sus funcionarios de salud, e independientemente de cuestiones personales o políticas, la evaluación del desempeño y de su espacio de trabajo, ha provocado que ministros del área de la salud en Perú, Ecuador, Brasil, Guatemala, Ruanda, Ucrania y Nueva Zelanda, entre muchos otros países, hayan sido relevados de su cargo por cuestiones relacionadas con evaluaciones de su desempeño en el trabajo (Mundt A, 2021). El ministro de salud de Ucrania, fue cesado por mala gestión de la pandemia y corrupción, al igual que en Perú, en donde han cambiado tres veces de ministros de salud en el periodo de la pandemia. Una inadecuada contabilidad entre el número de muertos reportados con respecto al exceso de mortalidad esperado en la estadística anual, además de declaraciones polémicas, llevó a renunciar al ministro de salud de Chile.

La evaluación en los espacios de trabajo (EET, "*work-based assessment*" o WBA por sus siglas en inglés), además de estar relacionada con el nivel más alto de la pirámide de Miller, el hacer (Miller, 1990), ayuda a entender la estrecha relación que existe entre la evaluación y el aprendizaje. Podríamos decir, que la educación es el fin último de la evaluación (Norcini, 2007). Lo anterior se ha explorado extensamente en el área de la educación médica, por lo que algunos de los ejemplos que se tratan en este capítulo están relacionados con la misma; sin embargo, la EET puede realizarse desde varias perspectivas y con diversos instrumentos, en áreas tan diferentes como el desempeño en trabajos altamente complejos, la influencia del ambiente en trabajadores con alto desempeño, el desempeño de deportistas, en donde el rendimiento se ve afectado por el trabajo en equipo, entre muchos otros (Govaerts, 2013).

En este capítulo se revisará el concepto de la EET, algunos ejemplos de cómo llevarla a cabo, de forma cuantitativa y cualitativa, la importancia de la retroalimentación, y se discutirán sus limitaciones, particularmente en cuanto a su evidencia de validez.

¿QUÉ ES LA EVALUACIÓN EN LOS ESPACIOS DE TRABAJO?

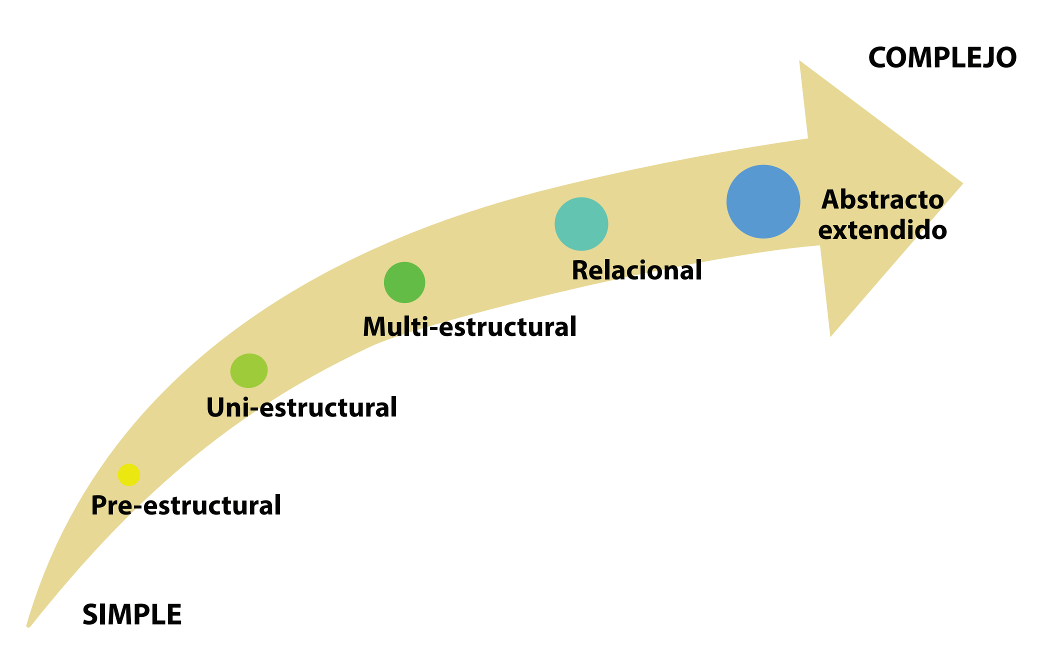
En los últimos 40 años se ha llevado a cabo una serie de cambios en la evaluación educativa que, históricamente, medía lo que podía medir, pero en ocasiones no lo realmente importante. Lo anterior se percibe en que actualmente se trata de evaluar lo que más importa y esto se logra con mayor precisión con las evaluaciones del desempeño (Crossley, 2012). El proceso de evaluación puede conceptualizarse como el conjunto de procedimientos que se utilizan para adquirir información de manera sistemática sobre algo o alguien, con la intención de efectuar juicios de valor sobre la toma de decisiones, y en donde queda implícito el aprendizaje (Luna de la Luz, 2020). La evaluación en los espacios de trabajo se refiere a un grupo de modalidades de evaluación que evalúa el desempeño de una persona, estudiante o profesionista, en el ambiente de trabajo real, lo anterior acompañado de una retroalimentación relevante, enfocada en la práctica reflexiva (Crossley, 2012). En lugar de enfocarse en instrumentos de medición específicos, la EET puede verse como un enfoque de la evaluación en espacios auténticos, que utiliza diversas herramientas y estrategias para determi-

nar el desempeño en el lugar de trabajo, pero no de forma aislada unas de las otras, sino de manera integradora.

La EET ha sido adoptada por diversas organizaciones, incluyendo escuelas de medicina y sistemas de salud, y es trascendental para medir el desempeño de médicos residentes durante su formación, graduación y certificación como especialistas. Por lo anterior, la EET puede ser aplicada en evaluaciones formativas, sumativas y de altas consecuencias (Yousuf, 2015). Este cambio de paradigma en evaluar lo que realmente importa, puede resumirse en los siguientes tres puntos:

1) El cambio de evaluar el conocimiento superficial y memorístico, hacia el de hacer pruebas sobre la comprensión, la construcción y la interpretación, se puede ver reflejado en la taxonomía propuesta por Bigg y Colloins (2010) llamada SOLO (siglas del término en inglés *Structure of Observed Learning Outcomes)* (Prakash, 2010), en donde se sitúa el desempeño basado en el conocimiento, de lo simple a lo complejo, en cinco niveles (Figura 1).

Figura 1. Taxonomía de SOLO
(Estructura de Resultados de Aprendizaje Observados)



Fuente: Elaboración propia.

2) El reconocimiento que las habilidades y destrezas son tan importantes como el conocimiento. Con base en la taxonomía propuesta por Bloom, se han desarrollado nuevos formatos para las evaluaciones en ambientes clínicos, particularmente en personas que cursan alguna especialidad (Miller, 2010). En la práctica, es posible decir que el conocimiento es necesario, pero no suficiente, para la comprensión, que esta es necesaria pero no suficiente para las habilidades (competencias), y que estas últimas son necesarias, pero no suficientes para la práctica deliberada y cotidiana.
3) Finalmente, desde la perspectiva psicométrica, tomando en cuenta que los evaluadores son subjetivos y la variabilidad en el desempeño por la especificidad de caso, se han desarrollado diversas "mini" pruebas, en diferentes formatos, como el examen clínico objetivo estructurado (ECOE) y el mini-CEX, o mini-ejercicio clínico de evaluación para enfrentar dichas problemáticas.

¿CÓMO REALIZAR LA EVALUACIÓN EN LOS ESPACIOS DE TRABAJO?

Son múltiples los métodos para la EET, que además se caracterizan por dar una retroalimentación adecuada y casi inmediata. De acuerdo con el método utilizado, pueden clasificarse de la siguiente forma, como se resume en la Tabla 1. En esta sección se describirán algunos de ellos.

Tabla 1. Ejemplos de estrategias de evaluación en los espacios de trabajo

Estrategia	Ejemplos
Observación directa	Mini-CEX: miniejercicios de evaluación en la clínica.
	Observación directa de habilidades en procedimientos quirúrgicos (DOPS).
	Habilidades no quirúrgicas para cirujanos (NOTSS).
	Evaluación de actividades clínicas confiables (durante la formación).
	Examen clínico objetivo estructurado (ECOE).
	Pase de visita práctico en el hospital.
Retroalimentación de múltiples fuentes	Evaluación de 360° (Mini-PAT).
	Revisión del desempeño médico (PAR).
	Evaluación durante un pase de visita en el hospital.

Auditoría y retroalimentación	Revisión por pares de un caso clínico y de los registros hospitalarios.
	Revisión de la mortalidad y morbilidad en una serie de pacientes.
	Auditoría de expedientes y evaluación de desenlaces.
	Revisión de expedientes y evaluación del razonamiento clínico (CSR).
	Discusión de casos clínicos.
	Expedientes de una serie de casos y factores de riesgo.
Simulación	Pacientes estandarizados en un ECOE.
	Simulación virtual.
	Simuladores de alta fidelidad.
Aprendizaje y portafolios (reflexión)	Portafolios de aprendizaje en línea (MAINPORT ePortfolio).
	Módulos del desarrollo curricular.

Abreviaturas de los términos en inglés: *Surgical DOPS (Directly Observed Procedural Skills), Non-Operative Technical Skills for Surgeons (NOTSS), Physician Achievement Review (PAR), 360 Degree Survey or MINI-PAT (Peer Assessment Tool), Chart Stimulated Recall (CSR), MAINPORT ePortfolio*: portafolio en línea para la certificación de médicos en Canadá (https://www.royalcollege.ca/rcsite/cpd/moc-program/about-mainport-eportfolio-e).

1) ***Miniejercicios de evaluación en la clínica (mini-CEX)***

El mini-CEX *(Mini-Clinical Evaluation Exercise)* se puede definir como un método de evaluación con observación directa de la práctica profesional, la evaluación es estructurada con un formulario y posteriormente se da retroalimentación al estudiante o residente. Fue desarrollado por John Norcini en 1995 y se basa en que las personas bajo formación profesional, inicialmente médicos residentes de medicina interna, eran observados por sus profesores durante diversos encuentros clínicos. En el trabajo original, cada aspecto del encuentro clínico era evaluado por los profesores con una escala de 9 puntos para el desempeño: 1 a 3 insatisfactorio, 4-6 satisfactorio y 7-9 superior. Cada encuentro con un paciente dura no más de 15 minutos, con unos 5-10 minutos de retroalimentación, por lo que el mini-CEX dura máximo 30 minutos (Norcini, 1995). Si bien originalmente fue desarrollado para residentes de medicina interna, actualmente se utiliza en diversas especialidades en el mundo. Dadas sus características, el mini-CEX puede ser útil para evaluar situaciones clínicas tan diversas como: dolor torácico, dificultad respiratoria, dolor abdominal, mareo, dolor de espalda y resolver problemas de diagnóstico clínico como artritis, hipertensión arterial sistémica o diabetes mellitus. Sus características más importantes son:

- Es una herramienta adecuada para la EET en cuanto a habilidades clínicas y otorga retroalimentación inmediata a la persona evaluada.

- Se basa en casos clínicos con pacientes reales y distintos observadores para cada situación clínica.
- Los casos clínicos son diferentes en cuanto a enfermedad y complejidad y son observados en diferentes ambientes.
- La información se recaba en una ficha estructurada.
- El tiempo promedio aconsejado es de 30 minutos, que abarca la observación y la retroalimentación.

Es un instrumento útil para evaluar en ambientes reales, las siguientes situaciones:

- Habilidades de comunicación al realizar una historia clínica.
- Habilidades de la exploración física.
- Profesionalismo.
- Razonamiento clínico.
- Habilidades de comunicación en general (dar malas noticias, pronóstico, tranquilizar a un enfermo).
- Organización de la información, eficiencia.

Se ha demostrado que el mini-CEX es un instrumento de evaluación confiable, con un coeficiente de reproducibilidad de 0.8, tanto para estudiantes de posgrado (residentes de especialidad) como de pregrado, en donde puede tomar algo más de tiempo (hasta 45 minutos). Puede ser utilizado en evaluaciones formativas y sumativas. Para la implementación del mini-CEX se sugiere seguir los siguientes pasos:

- La persona evaluada y la que realiza la evaluación conocen los objetivos de aprendizaje para un periodo concreto del plan de estudios. Se considera adecuado presentar al paciente, el observador se sitúa en una posición en la que no interfiera y no se distraiga.
- El observador evalúa la situación clínica, solo o con la coordinación de un tutor.
- El observador llena el formulario de evaluación y proporciona retroalimentación inmediata.

En el mini-CEX, la retroalimentación es inmediata, relacionada con la evaluación de una situación clínica que acaba de ocurrir. El proceso debe establecerse como un diálogo, en el que se promueve la autoevaluación y reflexión de lo que se hizo, con énfasis en los aspectos positivos, con refuerzo de ellos. Para lo anterior, se puede comenzar con preguntas como ¿qué considera que hizo bien? Por otra parte, los aspectos mejorables deben de indicarse de forma constructiva, con conclusiones y un plan de acción para el momento de educación que fue evaluado. Por todo lo anterior, el mini-CEX es un instrumento de EET fiable, representativo, con más ventajas que desventajas, que puede ser utilizado en diversos entornos para la evaluación en el pregrado y en el posgrado, para evaluaciones formativas y sumativas (Norcini J, 2003).

2) ***Tarjetas de encuentros clínicos***

Las tarjetas de encuentros clínicos (*CEC, Clinical Encounter Cards,* por sus siglas en inglés*)*, fueron desarrolladas en la Universidad de McMaster en Canadá (Hatala y Norman, 1999) y tienen algunas coincidencias con el mini-CEX. El propósito original fue evaluar situaciones clínicas con pacientes reales en médicos residentes de medicina interna, con evaluación del desempeño. En la tarjeta del artículo original (10x15 cm), se obtenía la siguiente información:

- Datos demográficos: fecha, nombre del estudiante y del evaluador.
- Datos del evaluador (adscrito o residente de mayor jerarquía).
- Hospital donde se realizó la evaluación.
- Principal encuentro evaluado: historia clínica, exploración física, comportamiento profesional, presentación del caso (verbal o escrita), razonamiento clínico (diagnóstico), razonamiento clínico (tratamiento).
- Evaluación del encuentro (escala de 1-5): 1: insatisfactorio, 2: debajo del nivel esperado, 3: en el nivel esperado, 4: arriba del nivel esperado, 5: sobresaliente.
- Comentarios sobre el desempeño.

En algunos casos se incluyen seis niveles de evaluación, en donde el nivel 5 es sobresaliente y el 6 corresponde al de alguien graduado (en pregrado o en especialidad). Esta EET es confiable, reproducible y tiene evidencia de validez para medir la competencia clínica después de 8 encuentros, lo que le da una reproducibilidad de al menos 0.8. El sistema, por dar una rápida retroalimentación, genera evaluaciones satisfactorias para los estudiantes.

3) ***Muestras de trabajo clínico***

Este método de EET, también desarrollado en Canadá, se basa en la recolección de información sobre un paciente, en diferentes dominios durante la admisión o la hospitalización, con tarjetas de evaluación para cada uno de estos momentos. Los dominios evaluados incluyen: habilidades de comunicación, interrogatorio y exploración física, agudeza clínica, comportamiento interpersonal (interconsultas con otros especialistas, equipo de trabajo, familiares), desarrollo profesional continuo y actividades de prevención y promoción de la salud (Turnbull, 2000).

La evaluación puede ser parecida a la de múltiples fuentes o de 360º, dado que en ella pueden participar personal de enfermería y los pacientes. En el siguiente ejemplo, se puede ver como la evaluación global, permite calificar los dominios señalados:

- Estrategias terapéuticas: saber qué hacer al momento de admisión del paciente.
- Habilidades de comunicación: con el paciente, la familia, el personal de enfermería, otros residentes e interconsultantes.
- Consulta con otros especialistas: solicitud adecuada de interconsultas.

- Manejo de los recursos: pedir los estudios necesarios, evitar pedir estudios *"para asegurarse"*.
- Plan de alta: explicado con claridad al paciente y a la familia o responsable.
- Relaciones interpersonales.
- Habilidades de colaboración.
- Habilidades de prevención y promoción de la salud: en personas adultas, saber su esquema de vacunación e indicar las vacunas faltantes, hacer exámenes de diagnóstico temprano de cáncer (citología vaginal, tacto rectal).
- Profesionalismo.

En esta EET, la escala es de cinco puntos, que va de no satisfactorio a excelente, para cada dominio. El método es fiable, con un coeficiente de 0.7 o más, después de siete evaluaciones. El método parece ser más útil para evaluaciones formativas que sumativas.

4) Encuentros al azar con pacientes

Forma parte de la enseñanza del pregrado en medicina y es a la que cotidianamente se ven expuestos los estudiantes de medicina en su formación clínica. Es una evaluación formativa, en la que grupos de 5 estudiantes participan en una enseñanza supervisada por un tutor, que guía el encuentro sobre aspectos de la entrevista clínica (elaboración de una historia clínica) y de la exploración física. El paciente no es conocido por los estudiantes, de ahí que el encuentro sea al azar o no planeado. Después de la presentación por el estudiante, es recomendable que el tutor demuestre la importancia del paciente estudiado, particularmente en aspectos de razonamiento clínico, diagnóstico y tratamiento. La retroalimentación puede darse en una escala de nueve puntos: desempeño debajo de lo esperado (1-3), desempeño adecuado (4-6), desempeño bueno (7-9). (Burch, 2006).

5) *Observación directa de habilidades de procedimientos* (*DOP*S)

Este método fue desarrollado en Inglaterra y se enfoca en evaluar las habilidades de procedimientos en los espacios de trabajo (DOPS: *direct observation of procedural skills,* por sus siglas en inglés). Se utiliza una escala similar a las previas, de seis puntos, en donde 1-2 corresponde a una evaluación por debajo de las expectativas, 3-4 al nivel de competencia esperado y 5-6 superior al esperado. El tiempo para la evaluación es de unos 20 minutos: 15 para la observación y 5 para la retroalimentación (Wragg. A, 2003).

En este caso se tiene una lista para la cual el evaluado debe mostrar su desempeño en maniobras tales como: intubación de la tráquea, colocación de una sonda intestinal, venopunción, administración de medicamentos intravenosos, administración de líquidos, toma de una gasometría arterial o habilidades quirúrgicas. Las evaluaciones se realizan en múltiples ocasiones por varios profesores. Si bien puede ser similar a la bitácora (personal o en línea), la diferencia es que en la bitácora clásica, el médico residente informa que ya ha cumplido la meta de ciertos procedimientos y la fecha en la que los realizó y con la información anterior se le considera competente, mientras que con la

DOPS, se asegura que la persona en formación es observada de forma directa, se le da retroalimentación inmediata y esto tiene implicaciones en cuanto a mejorar su desempeño en habilidades en los espacios de trabajo reales.

6) ***Discusión basada en casos***
Desarrollada de forma inicial por el *American Board of Emergency Medicine* en 1983, forma parte de la evaluación para la certificación de médicos de emergencias en Estados Unidos de América y en Inglaterra. En esta evaluación, el médico residente elige dos casos clínicos reales, con expediente, que son presentados al evaluador, quien elige uno de ellos y en 20 minutos otorga una calificación y retroalimentación de los siguientes aspectos: evaluación clínica, referencia del paciente, tratamiento, seguimiento, plan a futuro y profesionalismo, además de anotar de forma correcta la información en el expediente (Maatsch, 1983).

Este método cuenta con suficiente evidencia de validez e inclusive se utiliza en evaluaciones de altas consecuencias, como la certificación de especialistas. La evaluación correlaciona con la captura adecuada de la información en el expediente, exámenes orales, exámenes con pacientes estandarizados y recertificación de especialistas en medicina de urgencias con exámenes orales y escritos hasta 10 años después.

7) ***Evaluación con múltiples fuentes o de 360°***
A diferencia de los métodos antes mencionados, este tipo de evaluación tiene presencia en la industria y en los negocios, pues en ella intervienen figuras como los clientes, colaboradores, los colegas, los superiores y se realiza una autoevaluación. En medicina, puede aplicarse como una EET fiable cuando cuenta con 8 a 12 evaluadores, para exámenes formativos o sumativos, en el pregrado o en el posgrado y no corresponde a encuentros clínicos con un paciente, sino más bien al desempeño del estudiante en escenarios clínicos. La evaluación se puede realizar de la siguiente forma:

- El estudiante elige a 8 evaluadores, que pueden ser médicos especialistas, otros médicos residentes, estudiantes, personal de enfermería y otros profesionales de la salud o investigadores.
- Cada evaluador recibe un cuestionario estructurado que es igual al que usa el estudiante para la autoevaluación.
- Los cuestionarios son llenados y enviados para su procesamiento.
- Se evalúan categorías como: cuidado clínico, buenas prácticas clínicas, enseñanza y desarrollo de habilidades, comunicación con los pacientes y el equipo de salud, trabajo con colegas y una calificación general.
- La información, anónima, es capturada y enviada al profesor clínico, al menos dos veces por año, lo que le permite tomar acciones para mejorar el desempeño del estudiante.
-

Este tipo de evaluación puede distinguir entre médicos que están certificados o no por el consejo de especialidad correspondiente, más con relación en habilidades de conocimiento o técnico o con evaluaciones escritas, que con otras competencias como el profesionalismo (Whitehouse, 2002).

RETROALIMENTACIÓN EN LA EVALUACIÓN DE LOS ESPACIOS DE TRABAJO

La retroalimentación puede definirse como "la información otorgada a un estudiante por un profesor, un colega o uno mismo, en relación con el desempeño o el entendimiento" (Hewson et al., 1998). Dicha información le permite al estudiante confirmar, modificar, añadir, reestructurar y disminuir sesgos de la información en su mente, además de convertir la nueva información en conocimiento, metacognición y aumentar la confianza en el desempeño de tareas, razonamiento y estrategias.

El estudiante se enfoca en las siguientes preguntas: ¿hacia dónde voy?, ¿cómo voy?, ¿qué sigue? Una adecuada retroalimentación permite que el estudiante distinga el sesgo que se tiene de la realidad que percibe con lo que realmente sucede, una evaluación objetiva y en tiempo real, da perspectiva particularmente ante desempeños bajos y finalmente esta retroalimentación en tiempo real, permite que el estudiante modifique su conducta de forma más adecuada y rápida. Es importante que la retroalimentación se enfoque en la tarea que se evalúa, en el proceso y conductas, no en características de la persona que no sean modificables.

En la EET, es recomendable que la retroalimentación se realice bajo las siguientes condiciones, lo que favorece el aprendizaje:

- En un tiempo y espacio adecuados.
- Relacionarse con aspectos particulares del desempeño, no aspectos generales.
- Otorgarse en relación con decisiones o acciones, no con la interpretación de lo que se supone que el estudiante haría o tendría la intención de hacer.
- Debe darse en cantidades pequeñas y digeribles.
- Usar un lenguaje positivo, propositivo, no emitir juicios de valor o palabras negativas.

Existen múltiples recomendaciones de conductas positivas o negativas que deben considerarse durante la retroalimentación y que se resumen en la Tabla 2.

Tabla 2. Técnicas de retroalimentación que se sugiere evitar y adoptar

Técnicas que se deben evitar	Técnicas que se sugiere adoptar
Ambiente poco amigable, sin respeto, cerrado.	Ambiente respetuoso, no amenazante, amigable.
Hablar de pensamientos y sentimientos generales antes de la retroalimentación.	Dar una apreciación general, si es necesario, al terminar.
Establecerse como juez.	Usted va a enseñar, no a juzgar.
Enfocarse en la personalidad.	Enfocarse en el comportamiento.
Basarse en rumores o en la calificación de otros.	Basarse en los hechos observados.
Hablar sobre aspectos generales.	Hablar sobre aspectos específicos.
Dar demasiada o poca retroalimentación.	Dar la retroalimentación necesaria, en el tiempo establecido.
No dar ideas para mejorar.	Dar ideas para mejorar.
Dar demasiada o poca retroalimentación.	Basarse en metas definidas y conocidas.

Adaptada de Hewson y Little, 1998.

CONCLUSIONES

Así como "*enseñar es aprender dos veces*" (Joseph Joubert, Pensamientos, 1824), lo mismo podría aplicarse a la evaluación, en donde el estudiante y el profesor aprenden continuamente y mejoran su desempeño, particularmente en los espacios de trabajo. Las habilidades cuando son valoradas en un ambiente controlado, no necesariamente predicen su desempeño cotidiano, en la vida real. Conocer cómo una persona se desempeña en el día a día, implica que sea evaluado cuando está en su ambiente de trabajo. Esencialmente, realizar valoraciones en los espacios de trabajo, se puede traducir en beneficios tangibles para la organización, docentes, estudiantes y la sociedad en general.

REFERENCIAS

Burch, V. C., Seggie, J. L., Gary, N. E. (2006). Formative assessment promotes learning in undergraduate clinical clerkships. *South African Medical J ournal = Suid-Afrikaanse tydskrif vir geneeskunde, 96*(5), 430–433. http://www.samj.org.za/index.php/samj/article/download/1128/575

Crossley, J., Jolly, B. (2012). Making sense of work-based assessment: Ask the right questions, in the right way, about the right things, of the right people. *Medical Education, 46*(1), 28–37. https://doi.org/10.1111/j.1365-2923.2011.04166.x

Govaerts, M., van der Vleuten, C. P. (2013). Validity in work-based assessment: Expanding our horizons. Medical Education, 47(12), 1164–1174. https://doi.org/10.1111/medu.12289

Hatala, R., Norman, G. R. (1999). In-training evaluation during an internal medicine clerkship. Academic Medicine: *Journal of the Association of American Medical Colleges, 74*(10 Suppl), S118–S120. https://doi.org/10.1097/00001888-199910000-00059

Hewson, M. G., Little, M. L. (1998). Giving feedback in medical education: verification of recommended techniques. *Journal of General Internal Medicine, 13*(2), 111–116. https://doi.org/10.1046/j.1525-1497.1998.00027.x

Luna de la Luz, V., González-Flores, P. (2020). Transformaciones en educación médica: innovaciones en la evaluación de los aprendizajes y avances tecnológicos (parte 2). *Investigación en Educación Médica, 9*(34), 87–99. https://doi.org/10.22201/facmed.20075057e.2020.34.20220

Maatsch JL, Huang R, Downing S, Barker B. (1983). Predictive validity of medical specialist examinations. Final report for Grant HS 02038-04, National Center of Health Services Research. Office of Medical Education Research and Development, Michigan State University, East Lansing, MI.

Miller G. E. (1990). The assessment of clinical skills/competence/performance. Academic Medicine: *Journal of the Association of American Medical Colleges, 65*(9 Suppl), S63–S67. https://doi.org/10.1097/00001888-199009000-00045

Miller, A., & Archer, J. (2010). Impact of workplace-based assessment on doctors' education and performance: *A systematic review. BMJ (Online), 341*(7775), 710. https://doi.org/10.1136/bmj.c5064

Mundt, P.A. (2021). Assessing government responsibility for COVID-19 deaths. *The Lancet, 397* (10282), 1345. https://doi.org/10.1016/S0140-6736(21)00529-8

Norcini, J. J. (2003). ABC of learning and teaching in medicine: Work based assessment. *BMJ, 326*(7392), 753–755. https://doi.org/10.1136/bmj.326.7392.753

Norcini, J., Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. Medical Teacher, *29*(31), pp. 855–871. https://doi.org/10.1080/01421590701775453

Norcini, J. J., Blank, L. L., Arnold, G. K., Kimball, H. R. (1995). The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Annals of Internal Medicine, 123*(10), 795–799. https://doi.org/10.7326/0003-4819-123-10-199511150-00008

Norcini, J. J., Blank, L. L., Duffy, F. D., Fortna, G. S. (2003). The mini-CEX: a method for assessing clinical skills. *Annals of Internal Medicine, 138*(6), 476–481. https://doi.org/10.7326/0003-4819-138-6-200303180-00012

Prakash, E. S., Narayan, K. A., Sethuraman, K. R. (2010). Student perceptions regarding the usefulness of explicit discussion of "Structure of the Observed Learning Outcome" taxonomy. *Advances in physiology education, 34*(3), 145–149. https://doi.org/10.1152/advan.00026.2010

Turnbull, J., MacFadyen, J., Van Barneveld, C., Norman, G. (2000). Clinical work sampling. A new approach to the problem of in-training evaluation. *Journal of General Internal Medicine,* 15(8), 556–561. https://doi.org/10.1046/j.1525-1497.2000.06099.x

Yousuf Guraya S. (2015). Workplace-based Assessment; Applications and Educational Impact. *The Malaysian Journal of Medical Sciences: MJMS*, *22*(6), 5–10. http://www.mjms.usm.my/MJMS22062015/02MJMS22062015_ra.pdf

Whitehouse, A., Walzman, M., Wall, D. (2002). Pilot study of 360 degrees assessment of personal skills to inform record of in training assessments for senior house officers. *Hospital medicine* (London, England: 1998), *63*(3), 172–175. https://doi.org/10.12968/hosp.2002.63.3.2065

Wragg, A., Wade, W., Fuller, G., Cowan, G., Mills, P. (2003). Assessing the performance of specialist registrars. *Clinical medicine* (London, England), *3*(2), 131–134. https://doi.org/10.7861/clinmedicine.3-2-131

Capítulo 39

EXÁMENES PARA EL OTORGAMIENTO DE CERTIFICACIONES Y LICENCIAS A PROFESIONALES

Miguel A. Paniagua, Alex J. Mechaber, Amanda L. Clauser, David A. Johnson

RESUMEN E INTRODUCCIÓN

Fue recién en el transcurso del siglo pasado, en gran parte del mundo, que los campos como la medicina, el derecho, la ingeniería y la arquitectura realmente entraron a su fase de desarrollo en la cual pasaron a ser consideradas *profesiones* y no simplemente *ocupaciones*. Esta distinción es importante si queremos entender la función del otorgamiento de licencias a profesionales, como médicos y otros profesionales sanitarios y su justificación. Gran parte de nuestro enfoque en este capítulo se centra específicamente en la medicina como profesión y como un ejemplo de la estructura circundante que la respalda (por ejemplo, acreditación, otorgamiento de licencias, certificación y evaluación).

Las definiciones de lo que constituye una profesión con frecuencia consideran varios elementos clave. Estos incluyen definiciones del término "profesión" que se enfocan en un conjunto particular de conocimientos y/o habilidades; o educación y formación especializadas de carácter formal; o un compromiso de brindar, de manera desinteresada, asesoría o servicios objetivos a otras personas, al margen de las consideraciones enteramente financieras.

El sociólogo Eliot Freidson ofrece, tal vez, la distinción más importante para nuestros propósitos: la medicina como profesión se caracteriza por el alto grado de autonomía que la sociedad otorga a la naturaleza y al control del "trabajo" realizado por los médicos. Este grado de control se extiende desde la educación y formación de los futuros médicos hasta su incorporación formal en la profesión (es decir, el otorgamiento de una licencia) y los mecanismos que aseguran el cumplimiento de las normas profesionales y los estándares de la práctica profesional. El otorgamiento de tal libertad o autonomía es un reflejo de lo que muchos califican como un *acuerdo social* entre la profesión médica y los pacientes o el público a quienes esta brinda sus servicios.

DEFINICIÓN Y MEDICIÓN DE LA COMPETENCIA

El propósito de las evaluaciones para la obtención de credenciales, o exámenes usados para regular el acceso a profesiones y ocupaciones, es el de asegurar al público que los individuos que han obtenido licencias, certificaciones u otros endosos tengan los conocimientos, habilidades y capacidades (KSA, por sus siglas en inglés) para participar de forma segura en su profesión. Si bien, sentarse a rendir un examen es una parte altamente visible del proceso de obtención de credenciales, dicho proceso raramente depende de un examen solamente. Para incorporarse a la práctica profesional, cada persona podría necesitar completar un programa de educación o formación, demostrar sus KSA a través de la experiencia, supervisión o aprendizaje guiado de carácter profesional, y demostrar competencia por medio de una evaluación o serie de evaluaciones. Los sistemas de obtención de credenciales dependen de información suplementaria de múltiples fuentes para abordar la amplitud de las habilidades necesarias para ejercer como profesional.

Los exámenes de obtención de credenciales son muy importantes para las personas que los rinden, y "proporcionan al público, incluidos empleadores y agencias gubernamentales, un mecanismo confiable para identificar a los profesionales en ejercicio que han cumplido determinados estándares" (Standards, 2014. p.175). Como parte del proceso de obtención de credenciales, los exámenes pueden proporcionar una evaluación objetiva de los conocimientos relevantes para la práctica profesional y proporcionar una evaluación de los estudiantes en formación que sea independiente de las instituciones en las que reciben dicha formación. Dado el valor que los sistemas de obtención de credenciales dan a los resultados de los exámenes, es importante definir claramente el propósito y el uso o usos de los resultados de una evaluación antes de empezar a diseñar el examen y desarrollar el contenido del mismo. El propósito servirá como guía para las decisiones que posteriormente se tomen con respecto al diseño del examen, al desarrollo de las especificaciones del contenido y, finalmente, al desarrollo tanto del contenido del examen como de los procedimientos de obtención de puntajes.

Un estrecho vínculo entre una determinada práctica profesional y el propósito, el diseño, el contenido y las especificaciones del contenido de la evaluación es de vital importancia para que los puntajes del examen sean útiles para la toma de decisiones informadas en lo que respecta a la obtención de credenciales. El propósito del examen servirá como guía para el desarrollo de las especificaciones del contenido y, a su vez, los KSA usados para definir el alcance de la práctica profesional. Por ejemplo, si una evaluación se usa para establecer la competencia técnica o procedimental, es importante definir los procesos o procedimientos integrales a ser evaluados, así como las habilidades importantes en todos los procedimientos evaluados. Las especificaciones del contenido cumplen con diversos propósitos en el desarrollo de exámenes, entre ellos proporcionar información sobre el contenido evaluado a los examinados y a las partes interesadas, guiar a los encargados de desarrollar el examen en el desarrollo de modelos de equiparación y bloques de preguntas, y proporcionar detalles a los redactores de preguntas para el desarrollo del contenido.

Las especificaciones del examen suelen presentarse como listas o conjuntos de KSA. Dependiendo de la metodología utilizada para el desarrollo del examen, también pueden

presentarse como una matriz o constelación de competencias o procesos clave con los KSA relevantes categorizados dentro de cada una. Las categorías de alto nivel deben usarse como guía para el informe de puntajes y el muestreo de contenido, así como para la comunicación con respecto al alcance o detalle del material contenido en la evaluación. Los KSA detallados pueden desarrollarse a través de varios métodos; por lo general, esto empieza con un procedimiento conocido como análisis de la práctica profesional o del trabajo. Un estudio de análisis de la práctica profesional organiza información sobre lo que ciertos profesionales hacen dentro de su alcance definido de práctica profesional, así como los KSA que utilizan para desempeñar dichas tareas exitosamente. Esto genera datos que son tanto descriptivos como inferenciales con respecto a los KSA requeridos para desempeñar importantes tareas profesionales.

ANÁLISIS DE LA PRÁCTICA PROFESIONAL

Existen varios métodos bien establecidos para recopilar estos datos, algunos orientados a las tareas o enfocados en actividades profesionales, y otros orientados a las personas, o enfocados en características de la personalidad o las competencias requeridas para el éxito además de los KSA. Esta sección resume algunos enfoques comunes para el análisis de la práctica profesional en la obtención de credenciales. Si bien cada método es distinto, se espera que un estudio de análisis de la práctica profesional pueda utilizar múltiples métodos según lo requieran el propósito y el formato del examen. Por ejemplo, desarrollar KSA para un papel de proveedor médico requeriría un enfoque diferente para definir la práctica en comparación con el desarrollo de KSA para obtener una micro placa que indique competencia en el uso de software, y guiar el desarrollo de una prueba compuesta de preguntas de opción múltiple requeriría un enfoque diferente al de una prueba que evalúe habilidades procedimentales.

Los procedimientos orientados a las tareas se enfocan en la descripción de las obligaciones o actividades que las personas desempeñan (por ejemplo, "responder las inquietudes de los clientes a través de la comunicación escrita, presencial y telefónica"), o los tipos de problemas que se espera que los profesionales resuelvan, así como los métodos que utilizan para resolverlos exitosamente. Los métodos orientados a los procesos son en gran parte inferenciales y se enfocan en las cualidades personales requeridas para rendir con efectividad en una determinada práctica profesional. Estos pueden incluir KSA, así como rasgos y características como habilidades de comunicación técnica, empatía o pensamiento crítico efectivo en una situación de alta tensión. Dependiendo del propósito del examen, es importante elegir un método que reúna información orientada ya sea a las personas o a las tareas (o ambas) sobre los requisitos para ejercer la profesión de manera efectiva.

Los métodos se describen detalladamente en Clauser & Raymond. (2017) con ejemplos de escalas de calificación para aplicaciones orientadas tanto a las personas como a las tareas en Raymond. (2016). Los datos suelen recopilarse, según el propósito del examen, mediante encuestas y/o entrevistas a expertos en la materia que representan ampliamente a los miembros de la profesión. Este paso es de vital importancia, ya que las tareas de las personas que

ejercen en diferentes entornos o etapas de su carrera (o áreas geográficas o con una clientela diferente) pueden ser muy diferentes dependiendo de la profesión. Por ejemplo, las tareas de un profesional médico que ejerce principalmente en medicina general en un área urbana de gran tamaño probablemente sean bastante diferentes a las de una persona que trabaja principalmente en investigación médica o en un entorno rural.

Al lograr un equilibrio entre metodologías de recopilación de datos, como encuestas –que permiten recopilar una gran cantidad de datos relativamente superficiales sin imponer una carga excesiva en los encuestados con entrevistas más detalladas o trabajo en grupos focales– con, digamos, un pequeño grupo representativo que discuta o reaccione a los datos de la encuesta, podemos construir un esquema de KSA que represente los requisitos para ejercer la profesión exitosamente. Este enfoque híbrido también puede utilizarse para recopilar datos de tareas (a través de una encuesta, por ejemplo) así como datos de competencia (a través de grupos focales) a medida que los expertos en la materia reflexionan sobre las tareas desempeñadas en la práctica profesional, su necesidad para ejercer exitosamente y, posteriormente, cualesquiera competencias compartidas entre las diversas exigencias de la práctica profesional. Por ejemplo, una encuesta puede recopilar datos de profesionales sobre las tareas que desempeñan en diferentes entornos. En la medicina, los datos de una encuesta podrían indicar que los profesionales en diferentes entornos de la profesión desempeñan tareas como reunirse con pacientes, documentar [datos] adecuadamente, realizar exámenes físicos y pruebas de diagnóstico, interpretar y comunicar resultados, y desarrollar planes de tratamiento. Los grupos focales o paneles de expertos en la materia podrían indicar que algunas tareas, como la documentación, podrían ser frecuentes e importantes para mantener un consultorio, pero que no son esenciales y deberían excluirse del examen de obtención de credenciales. Luego de seleccionar uno o más métodos y de determinar la muestra de participantes y obtener sus respuestas a la encuesta, es importante planificar lo que pasará después, o la manera en que los resultados del análisis de la práctica profesional tomarán la forma de especificaciones del contenido y, por último, de un examen.

DESARROLLO DE LAS ESPECIFICACIONES DEL CONTENIDO

Históricamente, los esfuerzos de análisis de la práctica profesional eran impulsados por un grupo de expertos en la materia: posiblemente individuos veteranos o influyentes en el campo. Este enfoque abre las puertas a sesgos significativos en las especificaciones del contenido y los KSA, posiblemente dadas las exigencias específicas a la práctica profesional de los individuos seleccionados. Por ejemplo, si solo se usan educadores para guiar los esfuerzos de análisis de la práctica profesional, es posible que su trabajo sea un mejor reflejo de los objetivos del currículo o programa de formación que de las exigencias impuestas en personas ajenas a la educación en la práctica profesional clínica general. Abordar la posibilidad de sesgos es de vital importancia, y esta puede reducirse en gran medida a través de una recopilación de datos a gran escala (como una encuesta), un muestreo cuidadoso para asegurar que las

respuestas recibidas puedan generalizarse a diversos entornos de práctica profesional, especialidades y otras características únicas de una determinada profesión.

En lugar de usar a expertos en la materia únicamente para definir el alcance de la práctica profesional, incorporar un proceso de conexión impulsado por el juicio crítico puede proporcionar un contexto importante para las tareas, competencias u otros datos del análisis de la práctica profesional. Los expertos representativos en la materia pueden examinar los datos del análisis de la práctica profesional y proporcionar una perspectiva adicional con respecto a la importancia relativa de las tareas (por ejemplo, la documentación y los procedimientos de consentimiento informado probablemente no sean igual de importantes al enfocarse en temas como seguridad del paciente) o los comportamientos/competencias de carácter crítico para ejercer la profesión exitosamente. Incorporar el juicio crítico de esta manera proporciona un enlace vital entre los datos de la encuesta, las exigencias de la práctica profesional y las prácticas de desarrollo del contenido. Este enfoque puede incorporar conocimientos de expertos y al mismo tiempo limitar el sesgo, y si se completa de forma iterativa, puede incorporar a muchos grupos diferentes de profesionales y partes interesadas, lo que conlleva a una amplia capacidad de generalización de los resultados. Una vez que existe un modelo de práctica profesional, los grupos de expertos en la materia pueden desarrollar especificaciones del contenido que reflejen cada una de dichas áreas e identificar los KSA subyacentes para cada una de ellas.

Al igual que el enfoque descrito anteriormente, utilizado para conectar los datos del análisis de la práctica profesional, los datos de encuestas combinados con las contribuciones de expertos en la materia pueden utilizarse para establecer el valor relativo de las áreas de contenido que conforman el examen. Por ejemplo, si se determina que tres áreas de conocimiento y dos áreas de competencia son fundamentales para los profesionales, los expertos en la materia pueden usar datos de encuestas y sus propios conocimientos del campo para recomendar qué porcentaje del examen estará dedicado a la evaluación de cada área. Estos valores guiarán cuánto contenido de cada área se representará y/o qué proporción del día del examen se destinará a determinadas habilidades.

Para respaldar el papel de un examen en un programa de otorgamiento de credenciales y asegurar que los puntajes sean significativos para respaldar decisiones con respecto a la competencia de los examinados, es importante seguir un proceso defendible para establecer el alcance y la amplitud de la práctica profesional, identificar habilidades, comportamientos o competencias importantes a evaluar, y construir especificaciones del contenido que reflejen el propósito del examen. Existen muchos modelos establecidos, basados en importantes trabajos en el área de psicología organizacional industrial, que suelen usarse en combinación y que sirven para guiar un estudio de análisis de la práctica profesional. Incorporar el juicio crítico de expertos en la materia respalda la representación de KSA pertinentes a diversas áreas o contextos de la práctica profesional en un proceso de evaluación u otorgamiento de credenciales, y a su vez limita el sesgo y aumenta la relevancia del contenido del examen. En última instancia, estas prácticas respaldan un argumento de validez que, a su vez, respalda el uso de puntajes de examen para hacer inferencias con

respecto al grado de preparación de un individuo para recibir una credencial e incorporarse a la práctica profesional.

OTORGAMIENTO DE LICENCIAS EN LAS PROFESIONES

En 2021, el otorgamiento de licencias –no solo en medicina, sino en muchas profesiones– está firmemente arraigado en gran parte del mundo. Sin embargo, las implicaciones de una profesión regulada no son ajenas a la controversia. El economista Morris Kleiner ha resaltado una tendencia creciente hacia el otorgamiento de licencias ocupacionales en los Estados Unidos, Europa y China: una tendencia que aplica un modelo regulatorio a las ocupaciones mucho más allá de lo que alguna vez estuvo reservado para las artes doctas de la medicina y el derecho. Tan solo en los Estados Unidos, se estima que alrededor del 30% de la fuerza laboral actualmente ejerce en un campo regulado que requiere algún tipo de licencia o certificado.

Cabe destacar que un modelo de otorgamiento de licencias no es la única opción para una profesión regulada, como la medicina. En muchos países, la opción preferida es la de un modelo de registro en el cual la incorporación a la medicina se basa en la obtención de una credencial (es decir, el título de médico). Este modelo –utilizado en países como el Reino Unido, Canadá, Australia y Sudáfrica– también podría utilizar un examen obligatorio administrado por un organismo independiente, como el Consejo Médico General (General Medical Council) del Reino Unido.

La justificación mejor articulada para el otorgamiento de licencias yace en el concepto de *auditoría externa o independiente*. Este razonamiento empieza con un principio fundamental: que la misión del otorgamiento de licencias es proporcionar protección pública. Por lo tanto, las decisiones en torno al otorgamiento de licencias priorizan los intereses de los pacientes y del público por encima de los de la profesión y la educación médica. Por consiguiente, la función de auditoría externa o independiente del otorgamiento de licencias es asegurar, no asumir, que el graduado que posee el título requerido tiene los conocimientos y habilidades necesarios para incorporarse de manera segura a la práctica como profesional.

El hecho de que no puede asumirse la competencia ha sido reconocido incluso por algunos miembros de la comunidad educativa. Si usamos a la medicina como ejemplo, un comentario de 2019 en la Revista de Medicina de Nueva Inglaterra (*New England Journal of Medicine*) reconoció de manera explícita las múltiples presiones impuestas en los educadores médicos y sus instituciones: presiones como responsabilidad legal, deuda estudiantil y una proclividad fundamental, aunque comprensible, hacia los intereses del estudiante y de la facultad, pese a que en ocasiones estos podrían ser contrarios a los de los pacientes y el público. Tal como mencionaron los autores de este comentario, "… las facultades de medicina continúan graduando estudiantes que no deberían ser médicos".

El regulador médico tiene la mejor posición para ejercer un grado de rendición de cuentas de carácter profesional a través de sus papeles de otorgamiento de licencias, disciplina y elaboración de políticas. Esencialmente, construir sobre la base de salvaguardas integradas

al sistema, como acreditación de programas de educación médica, con el fin de asegurar que la profesión honre el acuerdo social.

El otorgamiento de licencias para ejercer la medicina (y también otras profesiones de la salud) desempeña un papel clave en este acuerdo social, al asegurar estándares y promover la rendición de cuentas de carácter profesional a través del espectro de la educación, la formación y la práctica profesional. Esencialmente, justifica la libertad que la sociedad ha otorgado a la profesión para dar forma al perfil de la práctica profesional, asegurando que los intereses del paciente/público no sean simplemente escuchados, sino que puedan prevalecer en caso de ser necesario.

Como asunto práctico, el otorgamiento de licencias es llevado a cabo por un grupo u organismo legalmente establecido. Nuevamente, si usamos a la medicina como ejemplo, esto puede ocurrir a nivel nacional (por ejemplo, el GMC) o a nivel más local (por ejemplo, consejos médicos estatales en los Estados Unidos). La composición de la membresía de estos organismos cuasi gubernamentales varía. En algunas instancias, la membresía está completamente compuesta de médicos; en otras instancias, los profesionales sanitarios y/o miembros del público aliados también pueden brindar sus servicios.

El otorgamiento de licencias es un elemento de las tres funciones tradicionalmente asignadas al organismo regulador de una profesión. Además de la revisión y confirmación de credenciales (es decir, el otorgamiento de licencias), estas autoridades tienen dos responsabilidades adicionales de manera rutinaria: (1) Examinación, que consiste en evaluar o examinar conocimientos antes de emitir la licencia, y (2) Disciplina, que consiste en tomar medidas o imponer sanciones en caso de faltas en temas de conducta, práctica o comportamiento profesionales u otras áreas.

En décadas más recientes, muchas autoridades encargadas del otorgamiento de licencias han dado cada vez más pasos proactivos en lo que se describe mejor como desarrollo de políticas. En la medicina, por ejemplo, estos podrían incluir identificar las mejores prácticas en áreas específicas de la medicina, establecer lineamientos en torno a procedimientos médicos, o adoptar y comunicar políticas oficiales específicas con respecto a problemas clave.

En lo que a exámenes se refiere, el papel de las autoridades encargadas del otorgamiento de licencias puede variar considerablemente. Algunas de estas autoridades siguen elaborando y administrando sus propios exámenes. Sin embargo, a medida que la ciencia de la evaluación ha ido evolucionando y fortaleciéndose en los últimos cincuenta años, las autoridades encargadas del otorgamiento de licencias han adoptado diferentes enfoques para incorporar estos conocimientos expertos. Algunas de estas autoridades han desarrollado experiencia interna en el desarrollo y la administración de su propio examen de modo que la función de evaluación está incorporada dentro del mismo organismo de registro u otorgamiento de licencias. Algunos ejemplos de este enfoque en la medicina incluyen el Consejo Médico General del Reino Unido, el Consejo Médico de Nueva Zelanda y el Consejo Médico Australiano.

Otros han externalizado su examen a una agencia nacional de evaluación. Algunos ejemplos en este caso incluyen al Consejo Médico de Canadá, que proporciona las partes 1 y 2

del examen de calificación a las autoridades encargadas del otorgamiento de licencias provinciales en Canadá, y los consejos médicos estatales de los Estados Unidos, que utilizan el Examen de Licencia Médica de los Estados Unidos (USMLE), un programa coauspiciado del NBME y de la Federación de Consejos Médicos Estatales. Adoptar este enfoque no significa que las autoridades encargadas del otorgamiento de licencias estén renunciando al control de estos exámenes. Las autoridades encargadas del otorgamiento de licencias médicas suelen estar involucradas en el desarrollo, la gestión y el establecimiento de políticas para estos exámenes.

EXÁMENES PARA EL OTORGAMIENTO DE LICENCIAS

Al margen del enfoque o de la profesión, el examen se convierte en un punto clave en la decisión de emitir una licencia. Antes de abordar el contenido y la estructura usuales de dichos exámenes, existen varios puntos fundamentales a tener en cuenta. En primer lugar, el examen para obtener una licencia cumple con un propósito fundamentalmente diferente al de gran parte de las evaluaciones llevadas a cabo en un entorno educativo, como es una facultad de medicina. En este caso, la distinción es entre una evaluación formativa y una sumativa. Dentro del contexto de un entorno educativo, la evaluación es a menudo formativa y su objetivo es evaluar y monitorear frecuentemente el desempeño de los estudiantes para proporcionar retroalimentación continua que corresponda al progreso a lo largo del currículo. La evaluación para obtener una licencia es completamente sumativa: se evalúan los conocimientos de un candidato a una licencia en un determinado momento una vez concluido un periodo formal de instrucción o formación.

El estándar que se aplica para el otorgamiento de una licencia en las profesiones sanitarias no es el dominio del contenido que está siendo evaluado, sino la demostración de un nivel de competencia mínimamente aceptable necesario para brindar tratamiento a los pacientes de forma segura. Esta distinción es importante. En muchos países, o incluso en la mayoría de ellos, la obtención de una licencia es el primer paso antes de la formación especializada. Por lo tanto, el examen para obtener una licencia suele abarcar un área más amplia de evaluación de contenidos que a su vez abarcan múltiples disciplinas, en comparación con las evaluaciones más enfocadas asociadas con los exámenes de certificación para la formación especializada.

En el hemisferio occidental, los orígenes de los exámenes para obtener una licencia se remontan a la tradición europea de debate escolástico. Este modelo tenía como núcleo el diálogo entre el estudiante y el profesorado, un ejercicio de preguntas y respuestas orales diseñado para poner a prueba la profundidad y amplitud de los conocimientos del discípulo. Si bien las primeras evaluaciones para el otorgamiento de licencias con frecuencia reflejaban esta tradición oral de interrogatorio entre examinador y examinado, hace ya mucho tiempo que estos exámenes pasaron a formatos escritos (por ejemplo, preguntas de opción múltiple, preguntas de respuesta extendida y otros formatos). La fortaleza del examen para la obtención de licencias en la era moderna está cimentada en un siglo de refinamiento en la ciencia de la medición.

Se han codificado más de sesenta años de desarrollo y progreso en la ciencia de la medición (consulte los Estándares para la Evaluación Educativa y Psicológica [*Standards for Educational and Psychological Testing*]).

Existen desafíos y vulnerabilidades en torno a los exámenes para la obtención de licencias que hoy en día son más evidentes que nunca. Algunos de estos pueden verse a un nivel operacional. La interrupción a nivel mundial desencadenada por los esfuerzos para primero mitigar y luego contener la propagación del SARS-CoV-2 dejó al descubierto debilidades estructurales y administrativas arraigadas en la entrega de muchos de estos exámenes. El distanciamiento social y las restricciones de viaje impusieron obstáculos significativos en la administración rutinaria de los exámenes para obtener licencias, tanto administrados de manera "interna" por las autoridades encargadas del otorgamiento de licencias, como a través de contratos con entidades nacionales de evaluación, que dependen de proveedores de servicios (por ejemplo, Prometric o Pearson VUE) como plataforma de entrega. Los formatos de examen que utilizan pacientes humanos estandarizados para una amplia evaluación de las habilidades clínicas y comunicativas probaron ser particularmente vulnerables. Las experiencias de las autoridades encargadas del otorgamiento de licencias en los Estados Unidos, Canadá, Australia y el Reino Unido reflejaron una amplia gama de ajustes necesarios en la entrega operativa de estos exámenes: desde la transición a una entrega completamente virtual (en línea) del examen de habilidades clínicas hasta la suspensión del examen.

Dejando de lado los desafíos únicos para la evaluación de competencias basadas en habilidades durante una pandemia, incluso la transición de la administración del examen en instalaciones físicas a plataformas en línea probó ser todo un reto: crear plataformas de entrega en línea (o adaptar plataformas existentes), asegurar suficiente apoyo de supervisores para una entrega del examen en tiempo real y la necesidad de una conexión ininterrumpida a internet impusieron obstáculos para las autoridades encargadas del otorgamiento de licencias y sus agencias de evaluación, quienes no estaban debidamente preparadas para adaptar rápidamente la entrega de su examen.

CERTIFICACIÓN Y MANTENIMIENTO DE LA CERTIFICACIÓN

La certificación en la profesión médica es un contrato social que asegura la competencia del médico en ejercicio. Tal como se expuso en párrafos anteriores, una licencia profesional es un requisito mínimo para ejercer una profesión, mientras que la obtención de una certificación (como en el caso de la medicina y otras profesiones de la salud) es un proceso riguroso (y específico de acuerdo la especialidad) que identifica a los médicos que ameritan la distinción de ser llamados "especialistas". Históricamente, las especialidades médicas han otorgado la certificación de un consejo a los médicos que han completado satisfactoriamente una formación de posgrado como residentes y que han aprobado una evaluación inicial basada en conocimientos.

El concepto de un consejo médico especializado data de 1908, cuando el Doctor Derrick Vail lo propuso en su discurso como presidente de la Academia Estadounidense de Oftalmo-

logía y Otorrinolaringología (American Academy of Ophthalmology and Otolaryngology). No fue sino hasta después de la Primera Guerra Mundial que los médicos empezaron a referirse a sí mismos como "especialistas". Una respuesta a esto fue el establecimiento de programas de formación de residentes como "un camino reconocido rumbo a la especialización". Las sociedades profesionales se desarrollaron poco después para apoyar el trabajo de los médicos que empezaron a estrechar el enfoque de la atención proporcionada dentro de sus áreas de medicina clínica. El primer consejo especializado, el Consejo Estadounidense de Evaluaciones Oftalmológicas (American Board for Ophthalmology Examinations), fue colegiado en 1917, y otros consejos especializados surgieron poco después. Estos consejos pretendían ser "la solución adecuada para la certificación de especialistas". Más allá de la emisión de certificados de cualificación, los defensores de los consejos los consideraban beneficiosos para: 1) elevar los estándares de la práctica profesional clínica en los nichos especializados; 2) educar al público y a otras personas sobre las crecientes capacidades de los especialistas; 3) proteger al público de los médicos en ejercicio no cualificados; 4) establecer requisitos para la educación y formación en medicina especializada; 5) desarrollar recursos educativos para la preparación de especialistas; y 6) proporcionar una supervisión de los procesos de evaluación para el otorgamiento de certificaciones especializadas.

Hoy en día, en los Estados Unidos, el Consejo Estadounidense de Especialidades Médicas (American Board of Medical Specialties, ABMS), está compuesto por 24 consejos de certificación que otorgan certificaciones a médicos de 40 especialidades médicas diferentes y 87 subespecialidades médicas. Tal como se indica en párrafos anteriores, estos consejos dependen de estándares psicométricos en el desarrollo de exámenes para evaluar los conocimientos, las habilidades y el juicio crítico de los médicos especializados según los estándares de práctica profesional establecidos por su profesión y para otorgar certificaciones del consejo a aquellos que cumplan con estos requisitos y que se mantengan actualizados con respecto a los mismos.

En los Estados Unidos, un candidato exitoso para la certificación en una especialidad puede obtener una certificación en la especialidad médica asociada al cumplir con los siguientes requisitos: 1) completar una experiencia a tiempo completo en un programa acreditado de residencias, por lo general por un mínimo de tres años, y 2) completar exitosamente una evaluación de conocimientos y juicio clínico en la disciplina médica designada. A la fecha, más del 85% de todos los médicos con licencia en los Estados Unidos son especialistas médicos certificados por un consejo, y el número de certificaciones para diferentes subespecialidades continúa en aumento, habiendo tenido lugar la aprobación más reciente en 2018.

Tal como inicialmente era el caso para el otorgamiento de licencias médicas, la principal motivación para promover la especialización surgió de la preocupación de la profesión con respecto a la calidad de la atención y su necesidad de asegurar la competencia continuada en una determinada especialidad médica. La diferencia yacía en que la certificación especializada fue concebida como una cualificación de mayor nivel otorgada por una sociedad profesional, y no por una entidad gubernamental (como un consejo médico estatal). Básicamente, el otorgamiento de licencias médicas satisface un requisito legal para incorporarse al ejercicio

médico, mientras que la certificación representa el logro de la mínima competencia profesional requerida para el ejercicio en el campo especializado designado.

Inicialmente, las decisiones con respecto a la competencia en una especialidad específica se toman sobre la base de años de cuidadosa observación y evaluación por parte del profesorado en programas de formación de residentes. El marco estructural creado para estas evaluaciones está basado en las competencias fundamentales determinadas por sociedades profesionales y consejos médicos especializados. El reconocimiento de los rápidos cambios en los patrones de conocimientos y prácticas a lo largo del tiempo dieron lugar al movimiento de mantenimiento de certificaciones. El Mantenimiento de certificación (Maintenance of certification, MOC) requiere una demostración de profesionalismo continuo, participación en aprendizaje de por vida y rendimiento satisfactorio en evaluaciones cognitivas y de desempeño en la práctica profesional.

A la fecha, el impacto y la evidencia de mejores resultados en los pacientes de médicos certificados por un consejo han sido mixtos. Casi la mitad de los más de 70 estudios llevados a cabo, incluidos algunos estudios basados en la población, han revelado una asociación entre la certificación por parte de un consejo y resultados directos del paciente y reclamaciones por mala praxis. Otros no han revelado ninguna asociación, pero la mayoría de estos estudios se enfocaron en la certificación inicial por parte de un consejo, mas no en el mantenimiento de la certificación. Algunos estudios más recientes han revelado que el mantenimiento de la certificación estuvo asociado de manera positiva con los puntajes de desempeño del médico.

Los médicos en ejercicio llevan tiempo analizando cuidadosamente los programas de mantenimiento de certificación y exigiendo ver pruebas de su efectividad y relevancia. Los médicos mencionan el oneroso compromiso de tiempo que implican estos programas, la carga financiera asociada con los mismos y la falta de uniformidad en el proceso de revalidación de certificaciones entre diferentes especialidades. Además, ha habido una creciente preocupación e insatisfacción con respecto al papel del MOC en la capacidad de un médico para ejercer y recibir credenciales en hospitales y sistemas de salud en los Estados Unidos. Esto, por ejemplo, ha resultado en la presentación de una petición para el retiro del MOC, firmada por más de 22,000 médicos especializados en medicina interna.

El MOC ha experimentado una revitalización significativa como resultado de esta oposición. En respuesta, el ABMS formó la Comisión de Vista al Futuro (Vision for Future Commission), a la que se le asignó la tarea de desarrollar recomendaciones para abordar las inquietudes. En su informe final para el ABMS, la Comisión señaló que "el énfasis en la certificación continua debe estar enfocado en la disponibilidad de información comisariada que ayude a los diplomados a brindar cuidados clínicos mejorados. A falta de un beneficio sustancial en lo que al aprendizaje se refiere, las evaluaciones tradicionales poco frecuentes de alta importancia, sin importar cuán sólidas sean en términos psicométricos, se consideran inadecuadas como una futura dirección para la certificación continua". A partir de esto, el MOC ha hecho una transición hacia una naturaleza más formativa.

Los médicos de los Estados Unidos se han vuelto más tolerantes en lo que respecta a los programas de MOC. Estos programas se enfocan en el aprendizaje de por vida con incentivos para revisar y mantener los conocimientos. Estas actividades están mejor diseñadas y estructuradas para las necesidades de un médico clínico ocupado. Muchos programas ofrecen a los médicos retroalimentación inmediata al responder preguntas. Varios consejos del ABMS han desarrollado innovadoras alternativas a las evaluaciones tradicionales. Por ejemplo, el Programa de Memorándum de MOCA (MOCA Minute Program), desarrollado por el Consejo Estadounidense de Anestesiología (American Board of Anesthesiology), y los MOCA-Peds relacionados, diseñados por el Consejo Estadounidense de Pediatría (American Board of Pediatrics), proporcionan a los médicos un número personalizado y específico de preguntas de opción múltiple en línea cada trimestre. Estos programas han conducido a una mayor aceptación del proceso de MOC para la revalidación de credenciales y certificaciones por parte de la comunidad médica y, en última instancia, al éxito del mismo.

Gran parte de la experiencia con la certificación en la profesión médica puede aplicarse a otros campos profesionales. Un ejemplo se encuentra en el campo de la ingeniería civil. En 2020, Hains et al. usaron lecciones aprendidas del proceso de certificación y otorgamiento de credenciales en la profesión médica para presentar recomendaciones para la implementación de un sistema de certificación especializada para ingenieros civiles. Recomendaron que la Sociedad Estadounidense de Ingenieros Civiles (American Society of Civil Engineers, ASCE) adoptara la estructura general del sistema médico de otorgamiento de credenciales, es decir, que el otorgamiento de licencias administrado por el estado funcionara como la autoridad legal para ejercer la ingeniería y que la certificación especializada fuera administrada por la profesión. Su intención era asegurar que los estándares para la especialización en la ingeniería civil estuvieran enfocados en la competencia más que en el dominio. Recomendaron que la certificación especializada pasara por una organización independiente similar al ABMS. Habiendo aprendido de la controversia en torno al MOC en la medicina, insistieron en que, al diseñar el sistema de certificación especializada, los costos y el tiempo de preparación deben ser razonables y los estándares deben ser lo suficientemente rigurosos como para agregar un valor mensurable a través de la mayor competencia de los profesionales certificados por un consejo.

DESAFÍOS Y ESTADOS FUTUROS

Las entregas de exámenes, tanto en línea como computarizados, seguirán enfrentando desafíos de seguridad en el futuro. Ambos formatos facilitan un acceso amigable para el examinado al examen para la obtención de una licencia más allá de las limitadas administraciones de exámenes, tradicionalmente ofrecidas solo 2 o 3 veces al año en el pasado, y nos enfrentaremos a desafíos similares en la certificación a medida que los modelos de evaluación formativa sean adoptados con mayor frecuencia a través de las diversas especialidades médicas. La creación de ventanas de evaluación (oportunidades para rendir un examen) más largas y/o frecuentes puede ser problemática en términos tanto de los límites del contenido como del

desarrollo del formulario del examen, así como las persistentes conversaciones en línea entre examinados. Estas últimas, a menudo realizadas de manera anónima o bajo pseudónimos e incluso en las redes sociales, imponen desafíos únicos cuando un subconjunto de examinados participa en grupos comunitarios cerrados, tableros de mensajería y aplicaciones donde las conversaciones van desde el intercambio de consejos para la preparación para el examen hasta el acto explícito de compartir contenido visto en un examen reciente. La ironía para las autoridades encargadas del otorgamiento de licencias y certificaciones es que la misma tecnología que les permite ofrecer administraciones de exámenes y entregas del contenido de exámenes con mayor flexibilidad también sirve para socavar la integridad de los exámenes a través de las acciones de un subconjunto de examinados.

El otro gran desafío que enfrenta la administración de exámenes de alta importancia sigue siendo la evaluación de competencias fundamentales más allá de los dominios puramente cognitivos, como los conocimientos. Podría decirse que esta área ha sido la mayor fortaleza del examen para la obtención de una licencia. En el siglo XXI, esta instantánea de los conocimientos mínimos requeridos tomada en un solo momento se siente cada vez más fuera de lugar con respecto a las realidades de la práctica profesional en un campo como la medicina, donde el conjunto de conocimientos científicos se expande de manera exponencial día a día, de modo que la habilidad crítica requerida en estos tiempos consiste menos en cuántos conocimientos pueden demostrarse en un determinado momento y más en asegurar la capacidad del médico para acceder, evaluar y aplicar adecuadamente los conocimientos en un contexto clínico. Los exámenes para la obtención de licencias y certificaciones deben reforzar su capacidad de evaluar este conjunto de habilidades. Evaluar estas competencias "más blandas" es de vital importancia para las autoridades encargadas del otorgamiento de licencias y certificaciones, que son plenamente conscientes de que los casos disciplinarios a los que se enfrentan rara vez tienen que ver con conocimientos insuficientes, sino con otras competencias clave (por ejemplo, comunicación, desempeño efectivo dentro de un entorno de sistema o equipo, comportamiento profesional).

CONCLUSIONES

El otorgamiento de licencias y certificaciones a individuos en diversas profesiones debe estar respaldado por un rigor psicométrico y una relevancia con respecto a las expectativas de un determinado campo de práctica profesional. Los desafíos relacionados con la seguridad del examen y la deshonestidad de los examinados son interminables a medida que la tecnología avanza. Debe abordarse en mayor medida la expansión del enfoque de las evaluaciones más allá de los conocimientos cognitivos a áreas igualmente críticas para el desempeño en una determinada profesión. El principio fundamental de estos procesos debe ser el de proteger al público a través de mecanismos que aseguren el cumplimiento de los estándares acordados según se espera de los individuos cualificados.

REFERENCIAS

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME) (2014). Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association.

Clauser, A.L. Raymond, M. (2017) Specifying the Content of Credentialing Examinations, in Davis-Becker, S. Buckendahl, C. (Eds.) Testing in the Professions.

Raymond, M. R. (2016). Job Analysis, Practice Analysis and the Content of Credentialing Tests, in S. Lane, M. R. Raymond, T. M. Haladyna (Eds.), Handbook of Test Development (2nd ed., pp. 144–164). New York: Routledge.

Kleiner, Morris M. (2015), *Guild-Ridden Labor Markets: The Curious Case of Occupational Licensing*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. https://doi.org/10.17848/9780880995023

Melnick, D.E. (2009) "Licensing Exams in North America: Is External Audit Valuable?" *Medical Teacher*. *31*(3):212-4. https://doi.org/10.1080/01421590902741163

Sally A. Santen, et. al. (2019) "Kicking the Can Down the Road—When Medical Schools Fail to Self-Regulate," *New England Journal of Medicine*, *381*:2287-2289

David Johnson, Humayun Chaudhry, *Medical Licensing and Discipline in America: A History of the Federation of State Medical Boards* (Lexington Books, 2012)

David Nichols: "Maintenance of Certification and the Challenge of Professionalism". Recuperado de https://www.abp.org/sites/abp/files/pdf/nichols_moc_and_the_challenge_of_professionalism.pdf

Gray, B., Vandergrift, J., Landon, B., Reschovsky, J., Lipner, R. (2018). Associations Between American Board of Internal Medicine Maintenance of Certification Status and Performance on a Set of Healthcare Effectiveness Data and Information Set (HEDIS) Process Measures. *Annals of Internal Medicine, 169*(2), 97-105.

Hains, D. B., Ressler, S. J., Lenox, T. A. (2020). *Lessons Learned from the Implementation of Board Certification in the Medical Profession*. 2020 ASEE Virtual Annual Conference Content Access,

Kraft, R., Davis, N., Nilsen, K. (2018). Maintenance of Certification Part IV in Residency Methods and Impact on Practice. *Family Medicine, 50*(8), 597-604.

Russo, A. M., & Weinberger, S. E. (2020). Maintenance of Certification: A Work in Progress. *Chest, 158*(1), 50-52.

Xu, T., Mehta, A., Park, A., Makary, M. A., Price, D. W. (2019). Association Between Board Certification, Maintenance of Certification, and Surgical Complications in the United States. *American Journal of Medical Quality, 34*(6), 545-552.

Capítulo 40

ASPECTOS ÉTICOS EN LA EVALUACIÓN DEL Y PARA EL APRENDIZAJE

Gabriela de la Cruz Flores

"La evaluación, además de ser un proceso técnico es un fenómeno moral.
Importa saber a qué valores sirve y a qué personas beneficia.
Importa mucho evaluar bien, pero importa más saber a qué causas sirve [...]"
Santos (2018, p. 13)

"Para pasar basta memorizar las diapositivas del profesor." "¡Profe en el examen venían temas que no vimos en clase!" "Solo la profesora sabe qué evalúa ¡Hay de ti si le preguntas!" "¡Son injustas las calificaciones! No hubo retroalimentación a lo largo del semestre." "¡Los exámenes no son de opción múltiple sino de confusión múltiple!" "Un porcentaje de la calificación de X materia consistió en ayudar al profesor con sus investigaciones aplicando cuestionarios ¡Acabé inventando los datos con tal de pasar!" "El semestre pasado tuvimos un profesor que se burlaba de nuestros errores y nos ponía apodos." "Ahora con la pandemia cuando toca examen todos debemos encender cámaras y micrófonos, si falla tu conexión ¡Ya perdiste!"

Es probable que algunas de las situaciones referidas en el párrafo anterior nos resulten familiares y, tal vez, habrá quienes se asombren al descubrir que detrás de ellas hay asuntos éticos relacionados con la evaluación del y para el aprendizaje. El propósito del presente capítulo es construir, junto con el lector, un visor que permita desde la ética de la evaluación, repensar derroteros que contribuyan al aprendizaje, autonomía y dignidad de nuestros estudiantes. Coincidiendo con Silva (2003, p. 84) "una visión ética de la evaluación nos hace levantar la mirada y realizar un examen crítico del quehacer profesional, en términos de captar elementos de nuestro actuar docente que reflejan modos de comprender y regular nuestra práctica."

Antes de continuar es importante distinguir tres conceptos que llegan a utilizarse como sinónimos: *evaluación*, *acreditación* y *promoción*. En general, se entiende por *evaluación* como el proceso sistemático a partir del cual se recaba y analiza información sobre el aprendizaje y orienta la toma de decisiones para ajustar y mejorar tanto los procesos de aprendizaje como la práctica docente. Por su parte, la *acreditación* se relaciona con los criterios y requisitos que permitan la aprobación –*o no*– de una asignatura. Mientras tanto, la *promoción* se refiere a las regulaciones para continuar con el siguiente ciclo-*semestre / año´*-escolar. Como puede apreciarse dichos conceptos se imbrican, pero poseen una naturaleza diferenciada.

Además, los sistemas de evaluación, promoción y acreditación se relacionan con procesos centrales en las trayectorias escolares del estudiantado como son el aprendizaje, la permanencia y el egreso exitoso (Baquero et al., 2012). Y de manera paradójica, también pueden devenir en *zonas de riesgo* dando lugar a la reprobación, la repitencia, la desafiliación, el bajo compromiso con los estudios y el abandono escolar, asuntos que impactan la continuidad escolar y con ello, al derecho a la educación. Por ello, trastocar los sistemas de evaluación, promoción y acreditación da en el blanco de problemáticas que los estudiantes llegan a afrontar solos, cargando a cuestas el estigma del fracaso escolar (Maddonni, 2014).

Desentrañar desde la ética dichos sistemas permite adentrarnos en el cuestionamiento de fondo sobre los propósitos y fines de la formación universitaria. Sin desconocer u omitir la triada en cuestión, en el presente capítulo centraremos la discusión sobre la ética de la evaluación del y para el aprendizaje.

La ética de la evaluación del y para el aprendizaje nos remite a la reflexión sobre el para qué de la misma. Las preguntas sobre el qué, cómo y cuándo evaluar –*si bien relevantes*– languidecen ante el para qué de la evaluación. ¿Evaluamos para calificar? ¿Evaluamos para clasificar y jerarquizar? ¿Evaluamos para seleccionar? ¿Evaluamos para promover? ¿Evaluamos para incluir? ¿Evaluamos para emancipar? Coincidiendo con Valencia y Vallejo (2015, p. 225) "la finalidad de la evaluación desde una perspectiva ética ha de ser la función formativa como un proceso permanente."

En el presente capítulo se asume el para qué de la evaluación como una de las principales sendas para aportar al aprendizaje y a la autonomía de los estudiantes a través del encuentro dialógico y el incesante reconocimiento de sus capacidades y potencialidades, abriendo espacios para el pensamiento crítico, la creatividad, la libertad y la transformación de aquellos como agentes capaces de aportar al bien común a través de su labor como profesionales. Además, se concuerda con Álvarez (2014, p. 51) cuando advierte que la evaluación se relaciona con la formación integral de las personas, por ello: "es sobre todo una cuestión ética […]. Los aspectos técnicos adquieren sentido […] cuando están guiados por principios éticos. Si entre los aspectos técnicos preocupa la objetividad, entre los éticos la preocupación se centra en la acción justa, ecuánime, equitativa."

El capítulo se organiza en tres grandes apartados. En el primero, se bosquejan algunas preguntas clave que pueden guiar la reflexión docente a fin de atisbar procesos y prácticas de evaluación signadas por la ética, que en última instancia representen vetas para asumir el encargo social que implica la formación de profesionales.

En el segundo apartado, se esbozan aproximaciones teóricas hacia la evaluación del y para el aprendizaje y delinean implicaciones éticas. Se destacan los referentes de la justicia social para problematizar en torno a la ética de la evaluación del y para el aprendizaje. A partir de ello, se reflexiona sobre el para qué de la evaluación trascendiendo aproximaciones instrumentales que han perpetuado y legitimado prácticas asimétricas sustentadas en el uso del poder y que con frecuencia tienden a excluir.

Finalmente, en el tercer apartado se analiza un cúmulo de situaciones reportadas por estudiantes universitarios donde está en juego la ética de la evaluación del y para el aprendizaje.

En cada situación se identifican principios éticos y valores fundamentales tales como la beneficencia, la no maleficencia, la equidad, la justicia, la autonomía, la imparcialidad, la confidencialidad, entre otros.

A lo largo del capítulo, el lector encontrará una serie de recursos y apoyos que tienen como propósito promover la reflexión docente sobre la ética de la evaluación del y para el aprendizaje.

ÉTICA Y EVALUACIÓN O CÓMO ASUMIMOS LA FORMACIÓN DE PROFESIONALES

Para García, Jover y Escaméz (2010, p. 17) "la meta social de la docencia consiste en la transmisión de la cultura y la formación de personas críticas". Desde dicho planteamiento se desatan varias interrogantes: ¿En qué consiste la transmisión de la cultura y a través de qué medios y recursos tiene lugar? ¿Cómo favorecemos la transmisión de la cultura, en particular, aquella relacionada con el hábitus profesional y el corpus teórico, metodológico, epistemológico y valoral? ¿Qué procesos contribuyen al pensamiento crítico, a la participación y se cede la responsabilidad a las próximas generaciones en aras de aportar a la sociedad basados en el conocimiento? En tiempos como los que transcurren, se vuelve irrenunciable avanzar en la discusión sobre cómo promover el pensamiento crítico, en tanto este:

> […] tiene que ver con el desarrollo de la racionalidad e implica que los estudiantes lleguen a comprender lo que hace que un razonamiento sea bueno, a mejorar sus habilidades para observar e inferir, generalizar, expresar hipótesis, concebir alternativas, evaluar afirmaciones, detectar problemas y percatarse de la acción apropiada […] también supone determinadas actitudes como la curiosidad intelectual, la objetividad, la flexibilidad, la honestidad y el respeto al punto de vista de los otros (García, Jover y Escaméz, 2010, p.19).

Es así como para alcanzar la meta social de la docencia, es imprescindible el encuentro dialógico (Cabra-Torres, 2010; Franco, 2008). El diálogo cobra mayor significado cuando asumimos que la transmisión de la cultura no se da en el vacío sino es resultado de la interconexión entre los conocimientos previos y los contextos de referencia de los participantes, es decir, implican reconstruir dicha cultura a través de procesos inter e intrasubjetivos. Desde la ética docente y coincidiendo con García, Jover y Escámez (2010, p.22), la deliberación es fundamental, ya que a través de esta: "no solo se exponen las propias posiciones, sino que se busca comprender, penetrar en profundidad, el fondo de los problemas, para buscar posturas comunes basadas en razones aceptables por las partes."

Lo expuesto hasta el momento conduce a varios asuntos clave en el campo de la ética de la evaluación del y para el aprendizaje. Si asumimos que la meta social de la docencia es la transmisión de la cultura ¿Qué implica dicha transmisión? y ¿En qué medida la evaluación contribuye a su alcance? Estas interrogantes generan implicaciones prácticas que redundan en la necesaria articulación entre los procesos de enseñanza, aprendizaje y evaluación. La transmisión de la cultura en la educación universitaria difícilmente puede comprenderse sino como un medio a través del cual se aporta al desarrollo integral de los estudiantes y al

entramado sociocultural para asumir las responsabilidades que como profesionales nos competen en cada campo del conocimiento. Es así como dicha transmisión se vuelve un espacio para discutir y aportar a problemáticas que atraviesan al quehacer profesional. Por ejemplo, en estos tiempos que corren, la transmisión de la cultura implica reconocer la diversidad, tomar consciencia de las desigualdades para asumir responsabilidades como profesionales, apelar al compromiso con los más desfavorecidos, a la construcción de sociedades justas y solidarias, así como velar por el bien común, incluyendo la salvaguarda del medio ambiente.

En otras palabras, la transmisión de la cultura es un vehículo que aporta a la transformación de los individuos como agentes que se construyen y reconstruyen a través de la mediación de docentes y pares en ámbitos profesionales específicos. Desde esta perspectiva, la ética de la evaluación del y para el aprendizaje, puede situarse en una disyuntiva: promover sistemas de evaluación restringidos a la simple verificación de información que reproduzca el *statu quo* de las profesiones con lo cual, siguiendo a Moreno (2011, p. 132), "se asegura y perpetúa un aprendizaje rutinario y memorístico" o alentar sistemas de evaluación que aporten al pensamiento crítico y divergente que, a la postre, favorezcan la construcción de alternativas ante realidades sociales convulsas e inciertas. Este último escenario implica generar entornos de aprendizaje que alienten la participación de los estudiantes y promuevan la retroalimentación constante de sus aprendizajes, al mismo tiempo que incentiven la motivación y curiosidad constante para seguir aprendiendo y de manera gradual, asuman mayor autonomía cuyo fruto sea el aprendizaje permanente.

Coincidiendo con Agudelo-Torres et al. (2018, p. 72) la evaluación no es "una mera respuesta por el conocimiento, sino más bien una oportunidad y una pregunta por el acto mismo de conocer." Con el propósito de alentar el análisis y la construcción de estrategias de evaluación que aporten al pensamiento crítico y divergente, en la Tabla 1 se ofrecen al lector una serie de preguntas clave y enlistan algunos atributos deseables de dichas estrategias. Se entiende por estrategias de evaluación como el conjunto de métodos, técnicas y recursos que utiliza el docente para valorar el aprendizaje de los estudiantes.

Tabla 1. Estrategias de evaluación del y para el aprendizaje

Tópicos de análisis	Descripción
Preguntas clave	**Qué estrategias de evaluación…**
	o ¿Favorecen el aprendizaje autorregulable y autónomo?
	o ¿Permiten valorar logros y alentar el aprendizaje?
	o ¿Estimulan la participación, el diálogo y la colaboración?
	o ¿Permiten contextualizar el aprendizaje?
	o ¿Reconocen la diversidad de los estudiantes y potencializan el aprendizaje?
	o ¿Fomentan la evaluación formativa?
	o ¿Motivan y alientan?
Atributos deseables	o Articulan la evaluación diagnóstica con los procesos de aprendizaje.
	o Especifican los aprendizajes a alcanzar (clarifican dominios y niveles de desempeño).
	o Promueven el aprendizaje de todos y compensan de manera propositiva las desventajas de origen.
	o Reconocen aprendizajes externos al ámbito escolar.
	o Rompen con estigmas y clasificaciones *per se*.
	o Retroalimentan y ofrecen seguimiento a los procesos de aprendizaje.
	o Incluyen instrumentos para el automonitoreo del aprendizaje (rúbricas, portafolios, bitácoras y diarios).
	o Modelan los procesos de autoevaluación y coevaluación.

Fuente: Elaboración propia.

Por otra parte, problemáticas como la reprobación, la repitencia, la desafiliación, el bajo compromiso con los estudios y el lastimoso abandono escolar, pueden redimensionarse desde la ética de la evaluación del y para el aprendizaje. Es común que dichas problemáticas se asuman bajo lógicas que privilegian una visión dicotómica de la realidad escolar, la cual divide, selecciona y clasifica a los estudiantes entre "buenos" y "malos" edificando culturas escolares jerárquicas y excluyentes (Perrenoud, 2010). Como ejemplo la siguiente anécdota. Hace tiempo un profesor muy reconocido en cierta Facultad al iniciar su curso rezaba el mismo estribillo: "El 10 es para Dios, el 9 para el autor del libro, el 8 para mi, el 7 para los excepcionalmente brillantes y para la mayoría están el 6 y el 5." Con ello, perpetuaba una visión jerár-

quica, que animaba la competencia más atroz entre los estudiantes por cumplir la auto profecía del docente.

No está de más subrayar las consecuencias de los procesos de evaluación en la vida de las personas. Al respecto McArthur (2019, p. 39) advierte "el vínculo entre la evaluación y la vida actual y futura de los estudiantes: su bienestar, autoestima, identidad y oportunidades sociales y económicas". Leathwood (citada en McArthur, 2019, p. 40) agrega que:

> La evaluación se utiliza para proporcionar fundamento y legitimidad a las estructuras sociales y a las relaciones de poder de las sociedades modernas, y para señalar el lugar que ocupa cada uno dentro de estas. […] Puede influir no solo en cómo nos vemos a nosotros mismos, sino también en nuestras relaciones sociales con los demás y en cómo los vemos.

De cara a dichas problemáticas, desde la ética de la evaluación del y para el aprendizaje, privilegiando el diálogo, el aprendizaje y la responsabilidad del encargo social que asumimos como docentes, cabría preguntarnos ¿Cómo podemos revertir las injusticias, la falta de equidad y arbitrariedades en las prácticas y procesos de evaluación que socavan el aprendizaje, la autoestima y las expectativas del estudiantado sobre si mismos y el mundo que les rodea? Es así como la ética de la evaluación debiera conducirnos a la reflexión sobre nuestra práctica pedagógica (Valencia y Vallejo, 2015). En la Tabla 2, se sintetizan algunas preguntas detonadoras para reflexionar al respecto.

Tabla 2. Preguntas detonadoras desde la ética en la evaluación del y para el aprendizaje ante indicadores que aquejan el aprendizaje, la permanencia y el egreso exitoso

Indicadores	Preguntas clave para la labor docente
Porcentaje de reprobación	¿Por qué reprueban mis estudiantes? ¿Cómo puedo prevenir la reprobación? ¿Cuáles son los puntos álgidos en el aprendizaje y qué tipo de apoyos/andamios puedo incorporar o modificar?
Porcentaje de repetidores y recursadores	¿Cómo incluyo a los estudiantes repetidores y recursadores? ¿Cómo compenso carencias formativas? ¿Cómo interactúo con mis estudiantes y qué expectativas tengo sobre ellos?
Resultados de aprendizaje	¿Cómo garantizo el aprendizaje de todos? ¿Cómo favorezco al aprendizaje a lo largo de la vida? ¿Cómo articulan los estudiantes su aprendizaje para solucionar problemas y contribuir al bien común?

Fuente: Elaboración propia.

APROXIMACIONES HACIA LA EVALUACIÓN DEL Y PARA EL APRENDIZAJE E IMPLICACIONES ÉTICAS

Grundy (1998) en un análisis amplio sobre las prácticas del currículo desde la teoría de los intereses constitutivos del conocimiento elaborada por Habermas, propone un entramado conceptual de alta valía para el campo de la evaluación. El *interés técnico* se basa en la necesidad de sobrevivir y reproducirse que tiene la especie humana, para ello, las acciones de las personas se orientan hacia el control y la gestión del medio. El *interés práctico* busca comprender el medio a fin de que el sujeto pueda interactuar con él, así la interpretación, la razón y el juicio de los sujetos son imprescindibles. El *interés emancipador* se orienta a un estado de autonomía. Según Grundy (1998, p.35): "La emancipación solo es posible en el acto de autorreflexión". Además, la autora citando a Habermas señala:

> La autorreflexión es a la vez intuición y emancipación, comprensión y liberación de dependencias dogmáticas. El dogmatismo que deshace la razón…es falsa consciencia: error y existencia no libre en particular. Sólo el yo que se aprehende a sí mismo…como sujeto que se pone a sí mismo logra la autonomía. El dogmático…vive disperso, como sujeto dependiente que no sólo está determinado por los objetos, sino que el mismo se hace cosa (p. 35).

La aproximación de la teoría de los intereses constitutivos del conocimiento de Habermas permite perfilar tres grandes aproximaciones sobre la evaluación que permean concepciones y prácticas al respecto. Desde el *interés técnico,* la evaluación se restringe al control, a través del cual se "valora el producto, y la evaluación, para que goce de autoridad y legitimidad, ha de adoptar la forma de medida" (Grundy, 1998, p. 61). Esta forma de concebir a la evaluación redunda en prácticas que clasifican, contrastan y segmentan a partir de la medición y que, como se ha señalado con anterioridad, contribuye a la edificación de culturas escolares jerárquicas, dividiendo entre los "buenos" *–los que se acercan al producto esperado–* y los "malos" *–aquellos que se alejan del producto esperado–*. Álvarez (2014, p. 52) sumando a lo expuesto, refiere que: "Desde el interés técnico se busca obsesivamente evaluar con bases científicas para garantizar el rigor de los métodos racionalmente planificados que permitan la discriminación por vía matemática." Para Franco (2019), lo anterior forma parte de prácticas heredadas de la razón instrumental que convierte a la evaluación en un *salvoconducto cultural*, donde es más importante "para algunos sectores de la sociedad, la promoción, el certificado y el valor de cambio de la calificación, que los procesos formativos." Esta perspectiva empata con los usos de la evaluación como instrumento de exclusión basada en principios de mérito y esfuerzo, donde los excluidos tienden a ser estudiantes que han sido expulsados de oportunidades sociales y educativas ampliando el campo de las desigualdades (Connell, 2006).

Sobre el mérito y el esfuerzo conviene hacer un par de precisiones. Cuando son resultado del goce de mayor capital cultural y social, ciernen desventajas hacia aquellos grupos menos favorecidos. Además, tienden a reforzarse con sistemas de evaluación que segmentan y clasifican. Esto no significa que se esquive el peso del esfuerzo, el espíritu por superarse y

el ímpetu por trascender ampliado al máximo las capacidades humanas. El problema estriba cuando los privilegios vulneran la condición humana pasado por encima de otros.

Por su parte, desde *el interés práctico* la evaluación se sitúa en juicios para valorar el alcance de los procesos de aprendizaje: "la evaluación significará elaborar juicios acerca de la medida en que el proceso y las prácticas desarrollados a través de la experiencia de aprendizaje favorecen el *bien* de los participantes" (Grundy, 1998, p.111). Como el lector puede observar, desde esta perspectiva la deliberación y la reflexión de los implicados son imprescindibles para estimar los alcances en el aprendizaje.

Por su parte, desde el *interés emancipador,* la evaluación implica liberarse de la opresión donde los participantes asumen una mirada crítica sobre los procesos de aprendizaje. Cabe señalar que desde esta perspectiva la participación, la colaboración y la simetría en las relaciones son concomitantes a los procesos de evaluación. Es así como la evaluación:

> [...] forma parte del riguroso proyecto del grupo de otorgar significados a las cosas. A través de los procesos de autorreflexión, es posible que los mismos grupos elaboren juicios sobre la medida en que su organización es indicativa de la ilustración y emancipación (Grundy, 1998, pp.177-178).

El interés emancipador abre vetas de alto calado para la ética de la evaluación del y para el aprendizaje. En primer lugar, reconoce que es desde la otredad que tienen cabida los procesos de autorreflexión, es decir, es en comunión con otros que reconocemos el alcance del aprendizaje y las transformaciones en el colectivo a través del diálogo (Segura, 2010). Es aquí donde los procesos de autoevaluación y coevaluación adquieren significado, en tanto animan la autorreflexión y reflexión conjunta. En segundo lugar, la simetría en las relaciones conlleva una suerte de bucle en donde la evaluación aporta no solo a los más neófitos sino a los más experimentados. Dicha simetría cobra cuerpo en los procesos de retroalimentación pues a través del diálogo y la orientación puntual, gradualmente se adquiere mayor dominio del campo –*disciplinar o profesional*– reduciendo las distancias y asimetrías entre los participantes. En palabras de Agudelo-Torres (2018, p. 76), la evaluación es "una evocación a preguntarnos por el aprendizaje de aquellos que interpretan y transforman [...] es una pregunta por el acto mismo de aprender; una invitación al "hablar-nos" que le es consustancial al sujeto y a la escuela en la que dicho sujeto cohabita."

En la Tabla 3 se ofrece una breve escala para que el lector pueda valorar la frecuencia con que promueve la retroalimentación en su práctica docente.

Tabla 3. Escala sobre retroalimentación en la práctica docente (breve exploración)

En una escala de 1 a 5, donde **1** representa **Nunca; 2 Casi nunca, 3 A veces; 4 Casi siempre y 5 Siempre,** valore la frecuencia con que realiza las siguientes acciones.						
		1	**2**	**3**	**4**	**5**
1.	Ofrezco información clara sobre las actividades a realizar.					
2.	Detallo el desempeño esperado en cada actividad ya sea de manera verbal o por escrito.					
3.	Expreso las principales cuestiones a mejorar en un lenguaje claro y accesible.					
4.	Sugiero cómo mejorar el aprendizaje de cada estudiante de acuerdo con su desempeño.					
5.	Planteo preguntas a los estudiantes que les permiten reflexionar sobre su aprendizaje.					
6.	Promuevo que los estudiantes construyan alternativas ante distintos problemas o situaciones.					
7.	Aprovecho las TIC para retroalimentar el aprendizaje de los estudiantes ya sea de manera sincrónica o asincrónica.					
8.	Puntualizo información que permite a los estudiantes mejorar la planificación de las actividades de aprendizaje.					
9.	Fomento que mis estudiantes reconozcan logros en su aprendizaje.					
10.	Planeo junto con los estudiantes alternativas para mejorar su aprendizaje.					
10-20 ¡Acércate a tus estudiantes! 21-30 ¡Vas por buen camino! 31-40 ¡Lo estás haciendo muy bien! 41-50 ¡Felicidades por tu labor!						

Fuente: Elaboración propia.

La evaluación como proyecto emancipatorio, conlleva adentrarse en los referentes de la justicia social. Dicha perspectiva está íntimamente relacionada con la equidad y, además, subraya las consecuencias de la evaluación en la vida actual y futura de los estudiantes (McArthur, 2019). En la educación universitaria la evaluación desde la justicia social cobra mayor significado por el impacto en la formación y en el desempeño profesional de las nuevas generaciones. Así que evaluar desde la justicia social, implica por ejemplo: promover el aprendizaje de todos y compensar de manera propositiva las desventajas de origen; garantizar aprendizajes indispensables para desplegar capacidades y dignificar la condición humana; romper con

estigmas y clasificaciones *per se*; reconocer aprendizajes externos al ámbito escolar; favorecer la autoevaluación y la coevaluación, como vías para la autorregulación y el aprendizaje a lo largo de la vida; transitar de prácticas de evaluación ligadas al uso del poder y la verticalidad a la democratización de los procesos.

En síntesis, los referentes teóricos expuestos sobre la evaluación como proyecto emancipatorio abren horizontes para la ética de la evaluación del y para el aprendizaje en tanto colocan en el centro el desarrollo pleno de los individuos y el papel de los docentes como gestores de autonomía y promotores del juicio crítico y el razonamiento. Al respecto, Valencia y Vallejo (2015) destacan la pertinencia de:

> […] vivir la evaluación como un proceso comprensivo de análisis del desempeño del estudiante, que implica el acompañamiento de quien enseña ligado al proceso formativo de quien aprende y la posibilidad de un perfeccionamiento constante [para ello] es necesario replantear las relaciones intersubjetivas que se entretejen entre el evaluado y el evaluador; para que estas favorezcan la vivencia del poder, más que como ejercicios de dominación, subyugación o sometimiento, como servicio y como una práctica democrática (p. 222).

LA ÉTICA EN LA EVALUACIÓN DEL Y PARA EL APRENDIZAJE: ANÁLISIS DE SITUACIONES

En este apartado a partir del análisis de situaciones descritas al inicio del presente capítulo, se analizan ciertas prácticas tácitas y asimiladas en nuestro quehacer cotidiano como docentes universitarios, donde con frecuencia llega a pasar desapercibido el uso de la evaluación como un mecanismo de poder, asimetría y exclusión. Y es que la evaluación en el aula suele moverse entre dos filos: como una vía que favorece el aprendizaje o como un instrumento de poder que mitiga cualquier ápice de autodeterminación. De ahí que, para potencializar la evaluación desde la reflexión ética resulta ineludible mirar tras bambalinas.

El hecho de que prácticas de poder embebidas en la evaluación pasen desapercibidas no es fortuito. En palabras de McArthur (2019, p. 41): "el dar por supuesto el estado de la evaluación y nuestra disposición a aceptar el estado actual de las cosas, son las actitudes que permiten que sus impactos dominantes y negativos en el mundo social continúen, en gran parte, pasando desapercibidos."

El carácter tácito de variadas prácticas de evaluación a menudo llega a asociarse con experiencias que los estudiantes describen como injustas en el ámbito escolar, por ejemplo: la arbitrariedad en la acreditación o no de los cursos; la indeterminación o la discrecionalidad en los criterios de evaluación; la ambigüedad en la ponderación de las actividades de aprendizaje; los usos de la evaluación asociados con el abuso de poder y la sumisión, mismos que provocan sentimientos de miedo, frustración, enojo, resentimiento y desesperanza. Siguiendo a Valencia y Vallejo (2015, pp. 220-221): "Con frecuencia los maestros recurren a la evaluación para mantener la disciplina, para infundir temor o para discriminar a sus estudiantes […] un mero instrumento de dominación, sometimiento, control, opresión y punición."

Por ello, resulta inaplazable discutir estos asuntos a fin de visibilizar desde la ética de la evaluación rupturas donde cada individuo encuentre nichos donde florezcan plenamente sus capacidades y potencialidades. ¿Acaso la evaluación, como proceso que forma, acompaña, retroalimenta, reconoce y motiva no tiene como fin aportar al bienestar y desarrollo integral de cada persona? Como docentes universitarios es importante crear espacios de discusión colegiados sobre el para qué de la evaluación. Esto puede dar lugar a la creación de comunidades que construyan alternativas y cuestionen aquello que se da como tácito e inamovible en las culturas escolares.

A fin de ilustrar situaciones frecuentes donde se pone en juego la ética en la evaluación del y para el aprendizaje se recuperan ejemplos de experiencias de estudiantes universitarios que denotan problemas éticos. Se precisa que la descripción de dichas experiencias son fruto de un proyecto de investigación en curso de naturaleza exploratoria donde se analizan las prácticas y procesos de evaluación en el ámbito universitario. Se invita al lector a valorar cada situación y en un ejercicio de introspección juzgue por qué considera –*o no*– que están comprometidos principios éticos en la evaluación del y para el aprendizaje.

Situación 1. Para pasar basta con memorizar las diapositivas del profesor

Líneas arriba se sostenía que una cualidad de la ética en la evaluación del y para el aprendizaje –desde el interés emancipador de Habermas– es el desarrollo de la autonomía de los estudiantes. La autonomía como principio ético implica la promoción de condiciones para que los individuos tomen decisiones basados en el juicio crítico y en la libertad. En la práctica docente, según García, Jover y Escámez (2010, p. 24) el objetivo central es promover la autonomía de los estudiantes, la cual "significa alcanzar la capacidad para regirse por el propio pensamiento y por las propias decisiones en los asuntos que a uno le conciernen."

Para Hortal (2002, p. 132) "Nada es verdaderamente humano si es impuesto a los hombres por otros hombres", esta cita cobra relevancia a la luz de la situación expuesta en tanto evaluaciones ancladas a la memorización inhiben el pensamiento crítico, la construcción del juicio propio y la asunción de responsabilidades como profesionales. Cuando el docente evalúa normas y rutinas, el estudiante pierde la "oportunidad de afrontar la evaluación en su dimensión formativa: alcanzar cuotas de autonomía intelectual para aceptar y relativizar los juicios externos [...]" (Cabra-Torres, 2010, p. 249).

Situación 2. "¡Profe, en el examen venían temas que no vimos en clase!"

No podemos evaluar lo que no enseñamos. Cuántas veces hemos escuchado la siguiente frase en un arranque de desesperación o de enojo: "¡Doy la clase por vista y el tema vendrá en el examen!" La coherencia de la práctica docente radica en armonizar nuestra vida y nuestra labor profesional. En la ética de la evaluación del y para el aprendizaje, la coherencia implica "hacer coincidir contenidos, competencias y estrategias [...]. Es exigir desde lo que me he exigido, es pedir sistematización desde lo que he sistematizado. Es recoger información que se ha dado con claridad y suficiente comprensión" (Valencia y Vallejo, 2015, p. 230).

Situación 3. "Solo la profesora sabe qué evalúa ¡Hay de ti si le preguntas!"

Santos (2018) a través de una actividad en clase ejemplifica el sentido de la evaluación como aprendizaje. Narra que pide a un voluntario participar con los ojos vendados en el juego del tiro al blanco, para ello, debe de tirar un gis al pizarrón donde previamente se ha trazado una gran diana con su blanco en el centro y anillos alrededor. Cada tiro es "evaluado" sobre una escala de 10 puntos y se comunica al voluntario el puntaje sin mayor explicación. Después de varios tiros, se pregunta al grupo ¿Aprenderá de este modo a hacerlo bien? ¿Servirá esta evaluación para el aprendizaje? Como el lector puede imaginar la respuesta es NO.

Una evaluación para el aprendizaje debe ser transparente y clarificar cuál es el puerto de llegada y cómo se emprenderá el viaje. Por ello, es necesario transitar de concepciones de la evaluación como una caja negra hacia una cultura de la transparencia donde los docentes faciliten la información sobre los criterios y estándares de la evaluación, así como las cualidades del desempeño por alcanzar a través de las diversas actividades de aprendizaje. Solo así los estudiantes podrán saber hacia dónde apuntar sus esfuerzos –*siguiendo la analogía del tiro al blanco*– y, además, la retroalimentación de pares y docentes tendrá mayor impacto. Para Parrondo (2004) brindar información también implica que los estudiantes conozcan y pongan en uso sus derechos para solicitar revisiones de exámenes y la protección de sus datos personales. La transparencia en la evaluación alienta la construcción de culturas escolares participativas y democráticas, lo cual redunda en la formación de ciudadanos que gozan y exigen el derecho a la información.

Situación 4. ¡Son injustas las calificaciones! No hubo retroalimentación a lo largo del semestre"

Para García, Jover y Escámez (2010), el principio de justicia en la práctica docente puede entenderse en dos sentidos. Por un lado, se refiere a la equidad entre los servicios prestados por el docente y el bien obtenido y, por otra parte, se refiere a la justicia social, donde se espera que los docentes sean solidarios con aquellos que más lo necesitan. Trasladando este principio a la evaluación del y para el aprendizaje, una evaluación justa será aquella que reconozca las diferencias de origen y de manera propositiva compense a fin de favorecer el aprendizaje de todos. Al respecto, López, Ordóñez y Rodríguez (2012) mencionan que "desarrollar una evaluación tecnicista, obsesionada con mediciones normalizadas hace que ésta sea indiferente e insensible a las diferencias económicas, sociales y culturales." En una sociedad tan desigual como la nuestra, evaluar con justicia implica generar procesos horizontales y dialógicos que favorezcan la retroalimentación continua de los procesos de aprendizaje. Para Agudelo-Torres et al. (2018, p. 72) "la retroalimentación de todo ejercicio valorativo […] ha de ser una oportunidad de aprendizaje, con tanta o quizá más importancia que el diseño y la posterior aplicación de un determinado instrumento evaluativo."

Situación 5. "Los exámenes no son de opción múltiple sino de confusión múltiple"

El papel de los exámenes en la evaluación del y para el aprendizaje ha dado lugar a álgidos debates en el campo (Parrondo, 2004). Uno de los puntos críticos sobre los mismos tiene que

ver con sus cualidades técnicas, es decir, elaborar un examen válido y confiable exige procedimientos que garanticen un diseño exitoso, para lo cual se espera que los docentes cuenten con la suficiente pericia para elaborar un examen que inspire confianza y veracidad. La construcción de exámenes válidos y confiables se robustece cuando se promueve la participación docente e incluso se suman a estudiantes a dicha empresa. Sobre la participación de estudiantes, Valencia y Vallejo (2015, p. 222) señalan que "no se trata de sustituir el "poder" del profesor por el del estudiante. Ambos deben y pueden razonar y justificar las decisiones evaluativas" lo cual contribuye a la corresponsabilidad en la evaluación.

Mención especial merecen los usos de los exámenes, ya que a partir de la discusión de los resultados y en particular sobre los errores puede afianzarse el aprendizaje. Cuando en el aula discutimos los "errores", abrimos cauces para reconocer posibles sesgos e incluso nociones en el proceso de construcción, con lo cual se alienta la reflexión y discusión colectiva sobre los conocimientos propios del campo profesional o disciplinar.

Situación 6. "Un porcentaje de la calificación de X materia consistió en ayudar al profesor con sus investigaciones aplicando cuestionarios. ¡Acabé inventando los datos con tal de pasar!"

El abuso de autoridad y el uso de los estudiantes para beneficio del profesorado, implica entre otras acciones, condicionar calificaciones. No hay que perder de vista la responsabilidad que se asume como docente y el encargo social encomendado. Condicionar calificaciones a cambio de actividades arbitrarias no solo vulnera el aprendizaje de los estudiantes, sino su dignidad ya que son concebidos como meros instrumentos al servicio de los docentes. Situación contraria es cuando los profesores invitan a estudiantes a sus proyectos de investigación y de manera libre participan asumiendo un rol activo que contribuye a los procesos de formación profesional, sin condicionar calificaciones o privilegios.

Situación 7. "El semestre pasado tuvimos un profesor que se burlaba de nuestros errores y nos ridiculizaba frente al grupo poniéndonos apodos"

García, Jover y Escámez (2010, p. 23) refieren que "el respeto a la dignidad de toda persona es el principio moral básico de la ética e implica respetar la conciencia, la intimidad y las características diferenciales de cada persona [*y rechazar*] toda clase de instrumentación o forma de violencia." En la ética de la evaluación del y para el aprendizaje el respeto a la dignidad de los estudiantes asume varios rostros. Por ejemplo, reconocer los diferentes ritmos de aprendizaje y ofrecer andamios que promuevan el máximo potencial de cada estudiante; respetar los diferentes puntos de vista y reconocer la diferencia como parte de una sociedad democrática; erradicar cualquier expresión de violencia producto de la evaluación, poniendo especial atención en aquellas situaciones donde de manera entreverada se violenta o discrimina por razones de género, condición sociocultural e incluso por discapacidad. Aquí se abre una veta de análisis de alta relevancia. Al desentrañar aquellas prácticas de evaluación que hemos asumido como "naturales" como es el caso de la ridiculización o burla en el aula, es probable que de manera tácita, la interseccionalidad de múl-

tiples variables genere desventajas para ciertos sectores universitarios, con lo cual se mina la confianza y la autoestima de los estudiantes condenándolos a ciclos de discriminación y exclusión. Esta situación es contraria al principio de beneficencia que debiese imperar en todo acto educativo.

Otro asunto que denota la situación es la falta de confidencialidad de la información recabada, más aún cuando esta se utiliza en prejuicio de los estudiantes.

Situación 8. "Ahora con la pandemia cuando toca examen todos debemos encender cámaras y micrófonos, si falla tu conexión ¡Ya perdiste!"

La evaluación del aprendizaje en la formación de profesionales en tiempos de COVID-19 ha sido uno de los procesos más críticos y de mayor controversia, no solo por las dificultades que representa evaluar con la mediación de las TIC en condiciones dispares, sino por el reto que representa la evaluación de habilidades y destrezas de la práctica profesional en espacios virtuales. En particular, la evaluación del y para el aprendizaje en carreras con un fuerte componente práctico las tensiones se han exacerbado y abierto debates sobre la urgencia de migrar hacia sistemas de evaluación que garanticen el aprendizaje y, por ende, la formación profesional.

La ética de la evaluación del y para el aprendizaje en la virtualidad implica generar sistemas que contribuyan con equidad al aprendizaje de todos y, además, que deleguen la responsabilidad y control a los estudiantes, solo así podremos migrar de culturas evaluativas basadas en la desconfianza a aquellas donde prime la responsabilidad, la coherencia y la honestidad como valores inherentes a los procesos de evaluación.

A MANERA DE CIERRE

La ética en la evaluación del y para el aprendizaje es un campo con nutrida discusión y donde incluso agencias nacionales e internacionales[1] han generado pautas para regular las prácticas y procesos comprometidos. Lo cual evidencia la relevancia y el impacto de los procesos de evaluación en la vida de las personas.

En el presente capítulo, la ética de la evaluación del y para el aprendizaje en la formación universitaria redunda en una aproximación que nos acerca al desarrollo de la autonomía y al fomento al pensamiento crítico y reflexivo. Desde la actividad docente, movilizar nuestra labor a fin de contribuir a dicha formación, implica tomar consciencia del encargo social que asumimos y que nos impulsa. Sin embargo, también es importante valorar aquellas prácticas encarnadas en la evaluación que asumimos como naturales y que lejos de aportar al aprendizaje vulneran la dignidad del estudiantado y acaban por mermar su confianza y perspectivas de futuro. Las situaciones expuestas solo pretenden poner en la mira la punta del iceberg e incitar la discusión colectiva a fin de construir y deconstruir una ética de la evaluación del y

1 Al respecto se sugiere revisar el sitio web: https://www.betterevaluation.org/en/rainbow_framework/manage/define_ethical_and_quality_evaluation_standards

para el aprendizaje en contextos universitarios. También habrá que mirar aquellas situaciones que denoten expresiones de generosidad, benevolencia y reciprocidad, a fin de andar por veredas que han contribuido a la formación universitaria.

La ética de la evaluación del y para el aprendizaje como se ha expresado a lo largo de este capítulo, requiere la creación de espacios participativos que fomenten el diálogo y potencien las capacidades de los individuos a fin de impactar en sus contextos de referencia. En aras de animar la creación de dichos espacios se propone:

a) Reflexionar junto con los estudiantes sobre las prácticas de evaluación del aprendizaje;
b) Hacer partícipes a los estudiantes del qué y para qué de la evaluación del aprendizaje;
c) Acompañar y modelar a los procesos de evaluación del aprendizaje, en este sentido no basta con emplear instrumentos novedosos si estos no se articulan con el aprendizaje;
d) Crear un clima de confianza en el aula y reivindicar la duda y el error como parte del aprendizaje.

Se espera que el lector al concluir la lectura de este capítulo pueda dar respuesta a la pregunta a qué causas sirve la evaluación y desde ahí pueda *tirar al blanco* en el fomento del aprendizaje de todos los estudiantes.

REFERENCIAS

Agudelo-Torres, J., Rojas-Restrepo, F., Ocampo-Ruiz, E. y Clavijo-Zapata, S. (2018). Sobre la evaluación escolar y su ética. *Información Tecnológica, 29*(5), 71-80. http://dx.doi.org/10.4067/S0718-07642018000500071

Álvarez, J.M. (2005). *Evaluar para conocer, examinar para excluir.* Madrid: Morata.

Baquero, R., Terigi, F., Toscano, A. G., Briscioli, B., y Sburlatti, S. (2012). La obligatoriedad de la educación secundaria: variaciones en los régimenes académicos. *Espacios en Blanco*, 22, 77-112.

Cabra-Torres, F. (2010). El diálogo como fundamento de comunicación ética en la evaluación. *Educación y educadores 13*(2), 239-252.

Connell, R. (2006). *Escuela y justicia social.* España: Morata.

Franco, E. (2019). Perspectiva ética de la evaluación educativa: reflexiones para deconstruir una representación de la evaluación como salvoconducto cultural. *Cuadernos Pedagógicos, 21*(27), 44-56.

Franco, Z. (2008). ¿Por qué la evaluación en el medio educativo debe ser ética para un desarrollo verdaderamente humano? *Revista Eleuthera, 2*(1), 144-164.

García, R. Jover, G. y Escámez, J. (2010). *Ética profesional docente.* España: Editorial Síntesis.

Grundy, S. (1998). *Producto o praxis del curriculum.* España: Morata.

Hortal, A. (2002). *Ética general de las profesiones.* España: Desclée.

López, J.M., Ordóñez, M.E. y Rodríguez, R. (2012). El papel de la ética en la evaluación educativa. *Unirevista.es,* 1, 87-95.

Maddonni, P. (2014). *El estigma del fracaso escolar: nuevos formatos para la inclusión y la democratización de la educación.* Argentina: Paidós.

McArthur, J. (2019). *La evaluación: una cuestión de justicia social. Perspectiva crítica y prácticas adecuadas*. España: Narcea.

Moreno, T. (2011) Consideraciones éticas en la evaluación educativa. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación, 9*(2): 130-144.

Ormart, E. (2004) La ética en la evaluación educativa. *Etic@net, 2*(3): 103- 112.

Parrondo, J.R. (2004). Aspectos éticos de la evaluación: impacto de la actividad evaluadora y códigos deontológicos de los examinadores. *Carabela,* 55, 31-43

Perrenoud, P. (2010). *La evaluación de los alumnos. De la producción de la excelencia a la regulación de los aprendizajes. Entre dos lógicas*. Argentina: Colihue.

Santos, M.A (2018). *La evaluación como aprendizaje. Cuando la flecha impacta en la diana.* España: Narcea.

Segura, M. (2007). La perspectiva ética de la evaluación de los aprendizajes desde un enfoque constructivista. *Actualidades Investigativas en Educación, 7*(1): 1-22. Recuperado de https://doi.org/10.15517/aie.v7i1.9242

Silva, M. (2003). Desafíos éticos de la evaluación educacional. *Revista Enfoques Educacionales, 5*(1): 81-87.

Valencia Rodríguez, W. A. y Vallejo Cardona, J. D. (2015). La evaluación educativa: más que una acción, una cuestión ética. R*evista Virtual Universidad Católica del Norte*, 45, 210-234. Recuperado de http://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/668/1199

Capítulo 41

EVALUACIÓN EDUCATIVA BASADA EN EVIDENCIAS

Melchor Sánchez Mendiola

"Artículo 3º. Todo individuo tiene derecho a recibir educación.
II. El criterio que orientará a esa educación se basará en los resultados del progreso científico, luchará contra la ignorancia y sus efectos, las servidumbres, los fanatismos y los prejuicios."
Constitución Política de los Estados Unidos Mexicanos

"La enseñanza necesita un ecosistema que soporte una práctica basada en evidencia. Necesitará mejores sistemas para diseminar más ampliamente los resultados de la investigación, pero también una mejor comprensión de la misma, para que los profesores puedan ser consumidores críticos de la evidencia."
Ben Goldacre

INTRODUCCIÓN

La práctica de la enseñanza en las universidades requiere informarse del cuerpo de literatura científica disponible en educación y ciencias sociales, para tender un puente entre el conocimiento original publicado y la docencia que ocurre en los espacios educativos físicos y virtuales. La experiencia personal del profesorado es fundamental e irreemplazable, pero es importante que esté informada por los hallazgos de la investigación sobre educación, para alcanzar un balance apropiado entre teoría y práctica y proveer un andamiaje académico a la actividad docente. La Educación Basada en Evidencias (EBE) ha surgido como un movimiento que propone utilizar la evidencia publicada para informar las decisiones y prácticas educativas en las universidades y escuelas (Dauphinee y Wood-Dauphinee, 2004; Goldacre, 2013; Hargreaves, 1996). Su diseminación e implementación han sido lentas por múltiples razones, por lo que es necesario reflexionar sobre sus avances y obstáculos en el escenario actual de la educación a nivel global (Hammersley, 2007; Hargreaves, 1996, 1997; Hammersley et al., 2014; McAleavy, 2015; McKnight y Morgan, 2020; Procter, 2015; Sánchez-Mendiola, 2012; Sharples, 2013; Simpson, 2019).

El uso de la evidencia académica publicada sobre actividades profesionales, como la docencia, se acompaña de retos conceptuales y prácticos. Muchos profesionistas basan sus decisiones y prácticas educativas principalmente en la experiencia, con limitado conocimiento de los avances en el campo de la educación y ciencias del aprendizaje, o de la amplia literatura científica disponible para informar sus actividades de enseñanza y evaluación (Camilli et al., 2020; CERI, 2007; Diery et al., 2020; Van der Vleuten, 1995, 2000; Wolf et al., 2001). Las universidades y centros académicos de investigación, idealmente deberían contar en su acervo bibliotecario con el material básico sobre cómo enseñar y evaluar en sus respectivas disciplinas (libros, revistas, bases de datos), entrenar a sus docentes en la búsqueda

de bases de datos sobre educación y en la evaluación crítica de artículos sobre esta temática, así como promover colaboración interdisciplinaria con académicos formados en el campo educativo. El enfoque de la EBE puede contribuir al desarrollo profesional del profesorado, integrando la docencia reflexiva como un elemento esencial para el aprendizaje significativo en los profesionistas que ejercen este complejo rol. Visualice los siguientes escenarios de evaluación:

- **¿Cuántas opciones deben tener las preguntas de opción múltiple?** Usted ha sido sujeto a exámenes de opción múltiple desde que era estudiante, y tiene sentimientos encontrados acerca de esta forma de evaluación. El responsable académico de su área en la universidad le pide 30 reactivos de opción múltiple para el examen final del curso de Farmacología. Le indica que sean reactivos independientes, con escenarios realistas, y seis opciones de respuesta para disminuir la posibilidad de que los estudiantes elijan la respuesta correcta por azar. Usted busca trabajos de investigación publicados en la base datos ERIC (https://eric.ed.gov) sobre el número de opciones ideal para exámenes de opción múltiple, y encuentra el meta-análisis de Michael Rodriguez sobre el tema (Rodriguez, 2005). Este trabajo documenta que tres opciones pueden ser suficientes para preguntas de opción múltiple.
- **¿Cuál es el mecanismo del impacto de las evaluaciones sumativas en estudiantes universitarios?** En su universidad se tiene la idea de que los exámenes determinan qué y cómo estudian los alumnos, lo que ocasiona que se preocupen más por los exámenes que por aprender lo que necesitan para ser buenos profesionistas. Usted le pregunta al encargado de exámenes si esto está demostrado y cuál es el mecanismo de estos efectos, él le responde que esto es lógico y no requiere investigarse. Realiza una búsqueda en Google Scholar y encuentra el artículo de Ciellers et al. (2010), en el que se revisa el tema y reporta un trabajo de investigación que provee evidencia sobre los mecanismos del impacto de la evaluación sumativa en los estudiantes.

En ambos escenarios se describe una situación en la que se requiere tomar decisiones educativas, así como buscar explicaciones fundamentadas a un fenómeno evaluativo que a veces es aceptado como axioma. Estas decisiones son susceptibles de ser informadas con la literatura científica sobre educación y evaluación. Los docentes involucrados tienen diversos grados de experiencia y preparación formal en los aspectos técnicos de la formación y evaluación de recursos humanos, y se encuentran insertos en estructuras organizacionales de diferentes niveles de complejidad que interactúan dinámicamente con su entorno social, económico y político. Realizar la búsqueda intencionada de la literatura académica para informar las acciones de evaluación educativa requiere actitud reflexiva ante las actividades cotidianas, capacidad de reconocer la ignorancia propia, iniciativa para buscar trabajos de investigación sobre el tema, así como habilidad para navegar en bases de datos y evaluar críticamente los artículos encontrados desde el punto de vista metodológico y de aplicación en su contexto. Este ciclo de tareas no es trivial, pero si conceptualizamos a la docencia como una actividad

profesional con cierto grado de sofisticación, informar nuestras tareas con evidencia de trabajos académicos debería ser la regla, más que la excepción.

El objetivo de este capítulo es ofrecer un panorama de EBE y su metodología, enfatizar la necesidad de orientar nuestras acciones educativas utilizando la información disponible en las publicaciones académicas, así como reflexionar sobre el estado actual de este campo y sus oportunidades para el futuro.

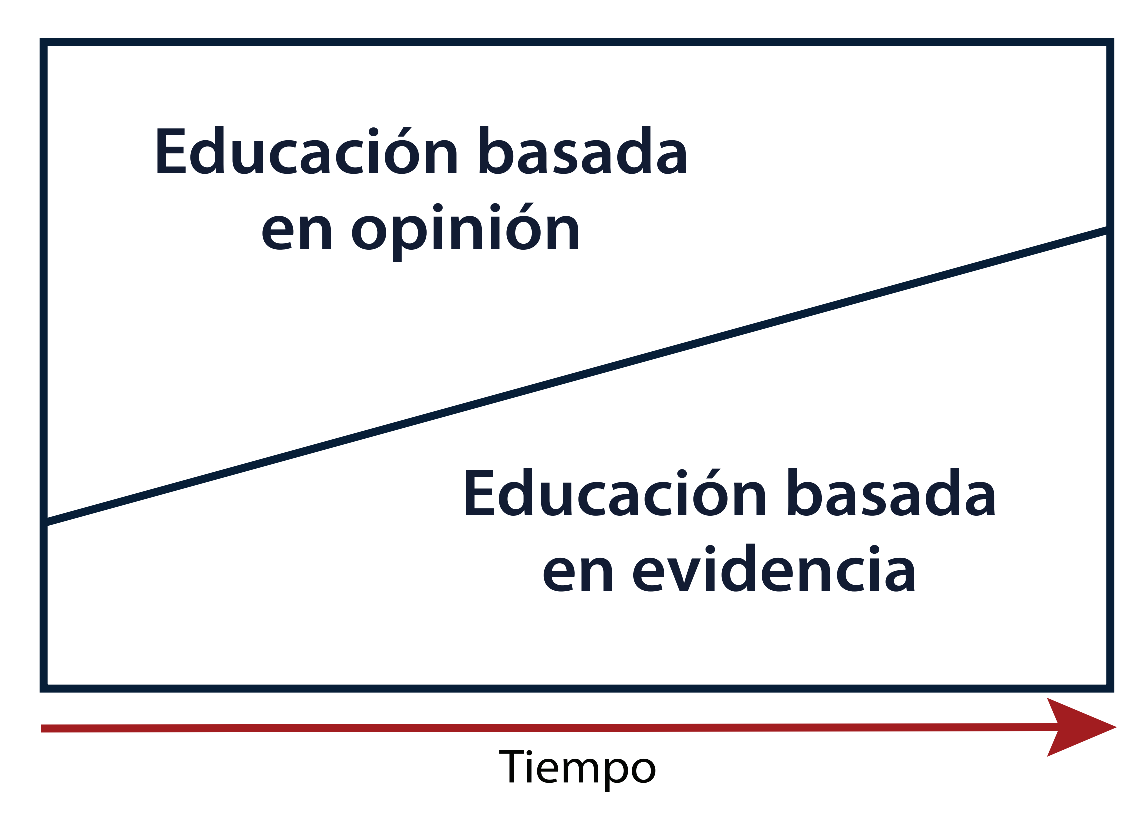
¿QUÉ ES LA EDUCACIÓN BASADA EN EVIDENCIAS (EBE)?

Los docentes tienen responsabilidades educativas, formales e informales, con pares, compañeros de trabajo y, principalmente, con estudiantes de diferentes niveles. Los profesores se inician en las labores educativas de diversas maneras, adoptando el rol docente con naturalidad y en ocasiones con poca preparación pedagógica y didáctica. Citando a George Miller: *"Es curioso que emprendamos tantas de nuestras responsabilidades más importantes sin preparación significativa. El matrimonio y la paternidad son probablemente los ejemplos más ubicuos, y existe poca esperanza de que estos estados humanos lleguen alguna vez a evolucionar racionalmente. La tarea de la educación médica, por otra parte, es aceptada deliberada y desapasionadamente, y sin embargo la preparación para tan influyente papel es igualmente frágil"* (Miller, 1980). A los profesionistas que se forman en un campo disciplinario, les es difícil captar la importancia de dominar los aspectos conceptuales y técnicos de la educación como campo de estudio durante su formación, ya sea porque no los visualizan en el currículo vivido, o porque los modelos de rol a que están expuestos los trivializan. Al iniciarse en docencia, la pasión por la disciplina propia, sea ingeniería, medicina, arquitectura, etcétera, permite creer que el dominio disciplinario es más que suficiente para enseñar, falacia que ha sido ampliamente analizada en la literatura académica (Centra, 1983; Hattie y Marsh, 1996; Hughes, 2005). Reflexione sobre las viñetas educativas con que inició el capítulo, ¿le suenan familiares?, ¿merecen los estudiantes universitarios un enfoque superficial a su proceso de aprendizaje?, ¿hay que conformarse con ejercer la docencia sin informarse cabalmente sobre los avances científicos de la misma? Parafraseando a Petersen (1999): *"Los profesionales del mañana necesitan educadores informados, no tutores amateurs"*.

En la década de los 90 inició un movimiento para enfrentar el problema del escaso uso de la evidencia publicada en la formación de médicos. La esencia de este movimiento fue descrita por Van der Vleuten, cuando ingresó al campo de la educación médica proveniente del área de la psicología: *"Me di cuenta que mis nuevos colegas -investigadores clínicos y biomédicos- tenían los mismos valores académicos que yo, lo cual me tranquilizó y me hizo sentir cómodo. Sin embargo, rápidamente me di cuenta de algo peculiar; las actitudes académicas del investigador aparentemente cambiaban cuando se discutían temas educativos. La evaluación crítica y el escrutinio científico eran súbitamente reemplazados por experiencias personales y creencias, y a veces por valores tradicionales y dogmas"* (van der Vleuten, 1995). Ian Hart de Canadá y Ronald Harden del Reino Unido propusieron el término BEME (*Best Evidence Medical Education*), lo difundieron en congresos y publicaciones y crearon la

BEME Collaboration para realizar revisiones sistemáticas sobre educación en ciencias de la salud (Harden et al., 1999; Harden, 2000; Hart, 2000). Harden y colaboradores argumentaron que la aplicación de la evidencia de investigación en educación se polarizó, con una aparente dicotomía en que la enseñanza era "basada en evidencia" o "basada en opinión", por lo que propuso que era conveniente conceptualizarlo como un continuo, en un extremo del espectro estaría la educación basada totalmente en opinión al no haber evidencia útil disponible, y en el otro extremo las decisiones basadas completamente en evidencia sólida publicada (Figura 1). Idealmente con el paso del tiempo, la comunidad incorporaría cada vez más la literatura publicada sobre el tema como un componente de sus acciones educativas.

Figura 1. La educación basada en evidencia puede representarse como un continuo en el que se combinan en diferentes proporciones las opiniones y la evidencia publicada



Adaptado de Harden et al., 1999.

La definición de EBE propuesta por Harden continúa vigente: "*La educación basada en evidencia es la implementación, por profesores en su práctica, de métodos y aproximaciones a la educación basados en la mejor evidencia disponible*" (Harden et al., 1999). La EBE requiere que el profesor actúe con profesionalismo y realice juicios de valor sobre el proceso educativo en que está inmerso, adaptándose al contexto en el que aplica la enseñanza. Para ello,

es necesario que el docente ejerza una práctica educativa reflexiva, identifique espacios de oportunidad en su práctica, genere preguntas susceptibles de ser contestadas por medio de trabajos de investigación publicados, evalúe críticamente la calidad de la información y la aplique de manera sensata en su contexto local. Adoptar la metodología de EBE no requiere que el docente sea un investigador en educación, más bien requiere una actitud, un *ethos* de docente reflexivo y pensamiento crítico permanente sobre el quehacer educativo, que responda a las necesidades del estudiantado.

En las últimas décadas ha ocurrido un intenso debate sobre el uso de la evidencia científica publicada en la toma de decisiones educativas, en el que algunos proponen que el estándar de investigación clínica (ensayos experimentales aleatorizados y meta-análisis) sea factor determinante de las acciones educativas, al tiempo que otros afirman que la educación tiene marcos epistemológicos que hacen imposible sujetarla a estos diseños metodológicos (Gorard et al., 2020; Gough, 2021; Guerra Ferrada et al., 2021; Hammersley, 1997; Hargreaves, 1996, 1997; La Velle, 2015; Munro, 2004). Al final del día, probablemente la situación idónea esté en algún lugar de este continuo, ya que si se considera a la docencia como una actividad profesional que debe alimentarse de los conocimientos de investigación original generados por la comunidad académica, es pertinente que el profesorado alimente su práctica con información metodológicamente válida. No se trata de que la experiencia, creatividad y libertad de cátedra del docente se vean determinadas o restringidas por artículos de investigación publicados, realizados en otros países y culturas, con recursos y realidades diferentes, sino que el uso de la evidencia publicada sea racional y contextualizado. En este sentido el uso de la palabra "basada en" es desafortunado, ya que da falsa impresión de rigidez y subordinación, cuando de lo que se trata es que los docentes sean consumidores educados de la literatura académica sobre educación en su disciplina, e informen sus acciones con la lectura crítica de esa literatura. No se trata de "recetas de cocina" educativas ni políticas rígidas prescriptivas de la autoridad, sino de integrar la sabiduría y experiencia profesional con la mejor evidencia externa producto de la investigación, para mejorar la práctica docente y el aprendizaje de los estudiantes. El término "práctica basada en evidencia" se utiliza cada vez más en el discurso público, tal vez un término más apropiado podría ser "práctica informada por la evidencia" (Chalmers, 2005). La palabra "evidencia" en el contexto de EBE, se refiere a publicaciones resultado de investigación, publicadas en la literatura académica con arbitraje por pares, que es uno de los pilares fundamentales para el avance de la ciencia a nivel global. Cuando se habla de evidencia en este capítulo no se refiere al significado tradicional o coloquial del término, como la evidencia de nuestros sentidos, la experiencia personal o grupal, las anécdotas, los informes institucionales, entre otros, sino a la evidencia publicada en la literatura académica.

Para el docente que se dedica primordialmente a su práctica profesional, la investigación original en educación puede verse como algo lejano, abstracto y poco aplicable. Si la comunidad de docentes con la que convive es de su misma área, hay poca motivación para que ocurra fertilización cruzada de ideas y retroalimentación de profesionales de la educación, que pudiera contribuir a mejorar su práctica docente. Movilizar el conocimiento de didáctica y

pedagogía no es trivial, es necesario ser proactivos y tender puentes entre las disciplinas para llegar a una educación y evaluación más integrales y sistémicas. Para ello se requiere esfuerzo y asegurar que el mundo académico y la práctica en el mundo real colaboren en sinergia con interacciones apropiadas. La meta final es relativamente clara: empoderar a los docentes con evidencias. En la Figura 2 se visualiza un esquema de la EBE, con los elementos del sistema y sus interacciones, en escenarios de contextos locales y globales.

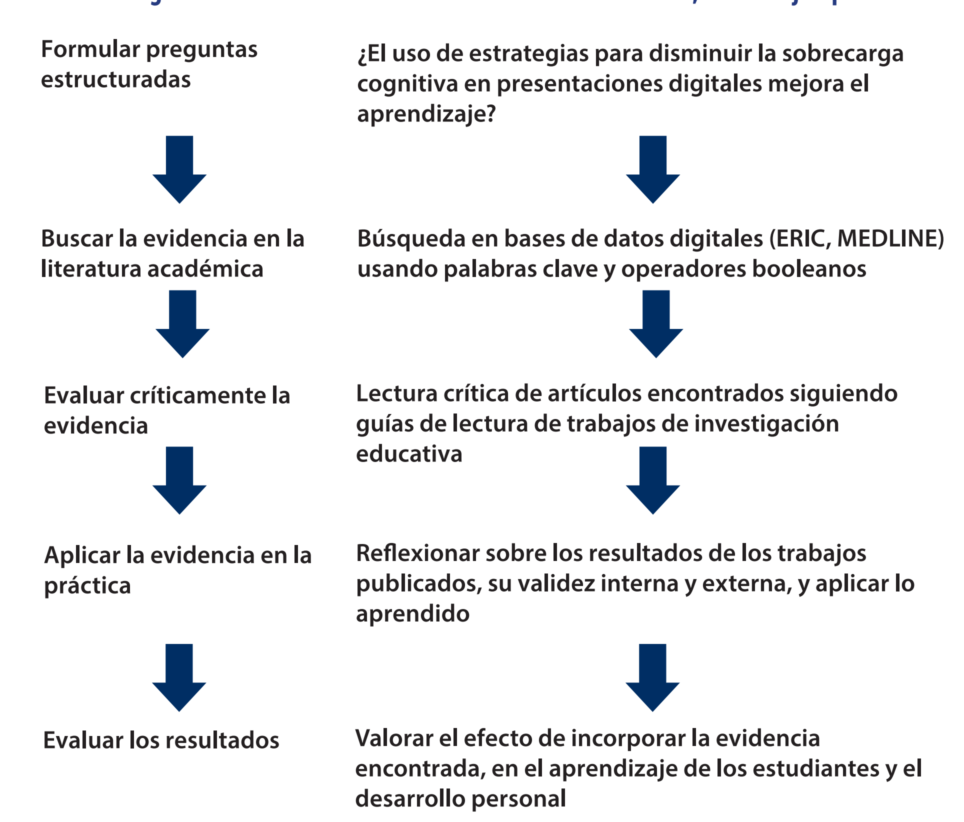
Figura 2. Modelo de la educación basada en evidencia en el que se representan sus elementos e interacciones

Adaptado de Hederich et al., 2014.

PASOS DE LA EBE

La secuencia de los pasos para incorporar la práctica reflexiva con el uso apropiado de la evidencia académica en la práctica docente se ejemplifica en la Figura 3.

Figura 3. Pasos de la educación basada en evidencias, con un ejemplo



A continuación, estos pasos se describen brevemente:

Generar preguntas estructuradas

El primer paso requiere generar una pregunta estructurada, relevante, contestable, que surja de la misma práctica docente. Para ello se requiere actitud de práctica reflexiva, la capacidad de reconocer vacíos de conocimiento propios y la intención de mejorar el desempeño. Si bien muchas veces las inquietudes que genera la práctica docente se quedan sin contestar, por exceso de trabajo y falta de tiempo, con frecuencia lo que se hace es preguntar a colegas, pedir opinión de expertos, o buscar superficialmente en la red. La esencia de la EBE implica buscar trabajos de investigación con rigor metodológico, en la literatura académica con arbitraje por pares, para obtener la mejor evidencia disponible. El acto docente y las interacciones con estudiantes generan preguntas, inquietudes y dudas, que requieren nutrirse con información. Un esquema utilizado en ciencias de la salud (PICO), puede ser de utilidad para estructurar las inquietudes docentes y encontrar con mayor eficacia resultados relevan-

tes en las bases de datos electrónicas (Eriksen y Frandsen, 2018). Este esquema se compone de cuatro elementos básicos que se pueden recordar con el acrónimo **PICO$_R$**:

- **P = Población o problema de interés.** Descripción de un escenario o problema educativo, que puede incluir variables como nivel educativo, área de conocimiento, entre otras, con cuidado de balancear brevedad con precisión.
- **I = Intervención o exposición.** Se describe la intervención educativa en estudio o la exposición a una situación de aprendizaje o evaluación.
- **C = Comparación.** Se describe una alternativa a la intervención propuesta inicialmente, como puede ser algo ya establecido o la no intervención. No todas las situaciones necesitan de una comparación.
- **O$_R$ = *"Outcome"* o Resultado.** Se recomienda describir un resultado educativo relevante para el estudiante, por ejemplo, aprendizaje, satisfacción, cambio de conducta, etc.

Los siguientes ejemplos usan las viñetas descritas al inicio del capítulo:

P = ¿En exámenes escritos con preguntas de opción múltiple…
I = …el uso de tres opciones de respuesta…
C = …comparado con cuatro o más opciones…
O = …mejora el desempeño de los reactivos manteniendo validez y confiabilidad?, ¿disminuye la inversión de tiempo de los elaboradores de reactivos?, ¿disminuye el tiempo de respuesta de los exámenes?

P = ¿En evaluaciones del aprendizaje en estudiantes universitarios…
I = …el uso de exámenes sumativos de alto impacto…
C = …comparado con evaluaciones formativas…
O = …afecta el aprendizaje de los estudiantes?, ¿afecta los métodos de estudio?, ¿cuál es el mecanismo de estos efectos?

Estructurar en forma de preguntas las lagunas de conocimiento que se encuentran en la práctica, incrementa la posibilidad de encontrar una respuesta, mejora el proceso de búsqueda en las bases de datos (la estructura de la pregunta sugiere términos de búsqueda), y mejora la comunicación con colegas, profesores y alumnos al verbalizar de manera puntual las inquietudes (Ericksen y Frandsen, 2018). Las preguntas constituyen un motivador poderoso para mejorar la calidad del desempeño docente, y ayudan a enfocar el poco tiempo disponible para leer literatura académica. Esta metodología ayuda a dirigir los esfuerzos de desarrollo profesional continuo a las necesidades individuales, y utilizarla en la práctica modela el rol de profesional reflexivo.

Buscar en las bases de datos digitales

Los profesionistas tienen sus preferencias de búsqueda de la literatura académica, generalmente dirigida a sus disciplinas. Sin embargo, a pesar de que hay abundante investigación

original sobre las ciencias del aprendizaje, didáctica y pedagogía, es poco frecuente que un docente busque explícitamente en las bases de datos digitales trabajos de investigación para mejorar su práctica docente y de evaluación (Delandshere, 2001; Diery et al., 2020; Harrison et al., 2017; Procter, 2015; Roberts et al., 2021). Habitualmente se enfoca en buscar evidencia del área disciplinaria propia o, en el mejor de los casos, artículos de educación sobre tareas específicas en su área, lo que excluye un caudal de información potencialmente relevante publicado en revistas especializadas de educación.

El paso para buscar la mejor y más reciente información requiere que el profesorado tenga conocimientos y habilidades de búsqueda de la literatura educativa, que habitualmente no se poseen al recibir responsabilidades docentes. Por ello, deben adquirirse por cuenta propia o buscar cursos y herramientas del tema en la red o en sus instituciones. El profesorado adscrito a una institución académica puede utilizar los recursos digitales disponibles en las bibliotecas, que actualmente se pueden usar a distancia desde cualquier lugar y horario, por medio de dispositivos digitales móviles y apps, lo que facilita el acceso a los artículos (Haig y Dozier, 2003a, 2003b; Pérez Puente, 2013). Las bibliotecas y bases de datos tienen recursos y motores de búsqueda para facilitar su uso, incluyendo tutoriales, bots para responder dudas, comunidades de aprendizaje y ayuda, así como catálogos digitales que facilitan el acceso, búsqueda y recuperación de la información sin costo para el usuario. Es obligación de las instituciones proveer el mayor acceso posible a estos recursos para la comunidad universitaria, y del profesorado exigir que estos recursos estén disponibles.

Para el profesorado que no tiene acceso a revistas que están detrás de "muros de pago" (*paywalls*), hay opciones como buscar en los navegadores gratuitos, usar revistas de acceso abierto, explorar sistemas donde se comparten versiones públicas de los documentos, o contactar directamente a los autores para que compartan sus artículos. Es responsabilidad personal del docente buscar opciones para informarse sobre los temas pertinentes a la enseñanza y evaluación, para mejorar su competencia como consumidor de información educativa. Nunca como ahora existen tantas opciones de búsqueda y recursos disponibles. En el Cuadro 1 se anotan algunos recursos con trabajos de investigación en educación.

Cuadro 1. Algunas herramientas para buscar información sobre trabajos de investigación en educación y evaluación educativa.

Nombre	Comentario
ERIC (Education Resources Information Center) https://eric.ed.gov	Base de datos del Departamento de Educación de EUA. Incluye literatura de investigación educativa, artículos de revistas, reportes, tesis, capítulos y libros .
What Works Clearinghouse (WWC) https://ies.ed.gov/ncee/wwc/	Centro de información sobre lo que funciona en educación, del Departamento de Educación de EUA.
Evidence4Impact https://www.evidence4impact.org.uk	Base de datos de intervenciones educativas del *Institute for Effective Education* (IEE) del Reino Unido.

MEDLINE https://pubmed.ncbi.nlm.nih.gov	Base de datos de la Biblioteca Nacional de Medicina de EUA, que incluye gran cantidad de artículos sobre educación.
Web of Science https://bit.ly/3GxYhu5	La base de datos de citas global más importante del mundo, requiere suscripción.
Google Académico https://scholar.google.com.mx	Motor de búsqueda de Google enfocado en la búsqueda de contenido y bibliografía científico-académica.
British Education Index https://www.ebsco.com/products/resea rch-databases/british-education-index	Índice de revistas educativas británicas, cubre políticas y administración educativas, evaluación, tecnología y necesidades especiales.
Scielo (*Scientific Electronic Library Online*) https://scielo.org/es/	Colecciones de revistas académicas de 17 países, publica revistas científicas y académicas en texto completo y de acceso abierto y gratuito.
IRESIE (Indice de Revistas de Educación Superior e Investigación Educativa) https://iresie.dgb.unam.mx/	Es un sistema de información especializado en educación iberoamericana del Instituto de Investigaciones sobre la Universidad y la Educación de la UNAM.
Evidence Based Education https://evidencebased.education	Organización británica dedicada a la educación basada en evidencia.
Evidence for ESSA https://www.evidenceforessa.org	Organización de EUA con información de programas que ayudan a que los estudiantes tengan éxito, preescolar a 12º grado .
Best Evidence Encyclopedia (BEE) https://bestevidence.org	Sitio creado por el *Center for Research and Reform in Education* (CRRE) de la Universidad Johns Hopkins, información sobre evidencia en educación, preescolar al 12º grado.
Best Evidence in Brief https://beibindex.wordpress.com	Boletín informativo del Center for Research and Reform in Education (CRRE) de la Universidad Johns Hopkins.
Evidence for learning https://evidenceforlearning.org.au	Organización de Australia para ayudar a usar la evidencia en educación.
Websites for Educators https://www.gse.harvard.edu/library/ed ucator-resources	Lista de sitios web útiles para educadores de la Escuela de Educación de la Universidad de Harvard.
Evidence Based Teachers Network https://ebtn.org.uk	Red británica de docentes basados en evidencia.
International Science and Evidence based Education (ISEE) Assessment https://mgiep.unesco.org/iseea	Iniciativa internacional de la UNESCO para educación y evaluación basadas en evidencia (Duraiappah et al., 2021).
Best Evidence Medical Education Collaboration https://www.bemecollaboration.org	Iniciativa internacional para revisiones sistemáticas en educación en profesiones de la salud.
Campbell Collaboration https://www.campbellcollaboration.org	Organización internacional para revisiones sistemáticas y síntesis de la evidencia en ciencias sociales.
Psycinfo https://www.apa.org/pubs/databases/p sycinfo	Índice de ciencias psicológicas, conductuales y sociales, de la American Psychological Association, con más de 5 millones de referencias.

Hay cada vez más libros, artículos, aplicaciones y sitios web sobre educación basada en evidencia en los diferentes niveles educativos, lo que facilita (al tiempo que complica, ya que hay que elegir y navegar en las diversas opciones) su aplicación en la práctica (Haig y Dozier, 2003a, 2003b; Pérez Puente, 2013). Se recomienda ser sistemático en el proceso, a través de tres pasos: definir y escribir el tema a buscar en la forma de una pregunta o hipótesis; identificar y expandir los conceptos esenciales; diseñar el alcance de la estrategia de búsqueda con criterios de inclusión y exclusión (Haig y Dozier, 2003).

Hay que recordar que las bases de datos y motores de búsqueda disponibles en la actualidad tienen un alto grado de sofisticación, y como cualquier nueva destreza, requieren tiempo, dedicación y motivación. Es importante recordar que no se trata de "hacer investigación" formal con un protocolo de búsqueda exhaustiva, se trata de obtener información vigente, válida y pertinente, sobre las tareas educativas y de evaluación a las que el docente se enfrenta. En ocasiones una búsqueda "rápida y sucia" en Google Scholar puede ser más que suficiente, en otras habrá que identificar las palabras clave, usar operadores booleanos y las opciones avanzadas de las herramientas. Si buscamos en Google Académico con términos simples "*multiple choice test*", "*options*" y "*meta-analysis*", el primer documento encontrado es el meta-análisis de Michael Rodriguez (2005) que responde la inquietud de la primera viñeta del capítulo; con la misma herramienta, si buscamos como términos simples "*summative assessment*" y "*mechanism of action*", el primer documento es el estudio de Cilliers et al. (2010), que responde a la segunda viñeta. Esto es un ejemplo de lo sencillo que puede ser encontrar literatura académica relevante a nuestras inquietudes de evaluación.

Analizar críticamente la evidencia

El siguiente paso, una vez que se recuperó un trabajo de investigación relevante a nuestra pregunta inicial, es la evaluación crítica del mismo. Esto no es sencillo ya que generalmente la formación de profesionistas no incluye el adiestramiento en la comprensión operativa de los conceptos más importantes de metodología científica, estadística, así como diseño de trabajos de investigación en educación y ciencias sociales (Newton et al., 2020; Procter, 2015). Es pertinente centrarse en los siguientes tres rubros:

a) **Análisis de la validez interna del estudio.** La metodología científica utilizada debe ser la apropiada para contestar el tipo de pregunta planteada en el trabajo de investigación. El diseño del estudio debe ser congruente con la pregunta de investigación, y de acuerdo al tipo de estudio se deben plantear una serie de preguntas sobre sus aspectos metodológicos, para verificar que los cumpla cabalmente y de esta manera confiar en los resultados del artículo. Por ejemplo, para contestar una pregunta sobre una intervención educativa en el que se haya realizado un estudio experimental con grupo control, deben satisfacerse criterios para documentar que se efectuó de una manera metodológicamente válida (McAleavy, 2015; Pigott et al., 2021). Ningún estudio es perfecto, la evidencia científica habitualmente se encuentra en tonos de gris,

y las herramientas utilizadas ayudan a obtener un perfil de la confiabilidad científica del estudio para tener una opinión educada sobre su validez.

b) **La magnitud de los resultados, su precisión y mecanismos de acción.** Tradicionalmente los docentes tienen ideas generales sobre la eficacia de los métodos de enseñanza y evaluación, pero no de la magnitud de sus efectos en términos cuantitativos y cualitativos. Se sabe que el aprendizaje activo es mejor que el pasivo, pero es poco usual que se tenga una idea clara de la magnitud de la diferencia o del cómo funciona. La EBE propone que se ponga énfasis en identificar la magnitud de los resultados de manera cuantitativa y cualitativa, por ejemplo, su impacto en el aprendizaje profundo de los estudiantes, la satisfacción de los docentes, o los mecanismos psicológicos del proceso.

c) **La validez externa o aplicabilidad del estudio en la práctica.** Este apartado es complejo, ya que los trabajos de investigación publicados habitualmente se realizan en poblaciones homogéneas, en contextos particulares que son difíciles de replicar en otros escenarios socioeconómicos. Lo anterior hace difícil usar los resultados de un estudio realizado en Estados Unidos, publicado en *Educational Researcher*, en una población de estudiantes que no refleja la realidad de México, y con ello extrapolar la intervención a nuestros estudiantes, con recursos limitados, en contextos muy distintos. Para aplicar la evidencia publicada, es importante que el docente use su pericia para identificar las necesidades de sus estudiantes relevantes para el aprendizaje, integrarlas con la información revisada y su experiencia personal, para así definir la mejor manera de enseñar y evaluar en el contexto local.

Es frecuente que el profesor carezca de los conocimientos y habilidades necesarios para analizar, desde el punto de vista metodológico, los trabajos publicados. Es recomendable tomar cursos presenciales o en línea para incrementar sus habilidades sobre el tema:

- https://www.edx.org/es/course/an-introduction-to-evidence-based-undergraduate-stem-teaching-6
- https://www.coursera.org/lecture/teachingscience/evidence-based-science-teaching-ew0YA
- https://www.mooc-list.com/course/using-technology-evidence-based-teaching-and-learning-futurelearn
- https://cursos.iadb.org/es/indes/what-works-education-evidence-based-education-policies.
- En la Universidad Nacional Autónoma de México (UNAM) se realizan cursos presenciales y en línea sobre esta temática (https://cuaieed.unam.mx), como han reportado Campillo y Benavides Lara y cols. (2019, 2021).

Existen diversos instrumentos para evaluar trabajos de investigación en educación, que pueden ser de utilidad como herramientas para la evaluación crítica sistematizada de los mismos (Cook y Reed, 2015; Education Group for Guidelines on Evaluation, 1999; Hannes et al., 2010; Morrison et al., 1999; Sánchez-Meca y Botella, 2010; Zisblatt et al., 2019). Se han

publicado las metodologías de evaluación de trabajos de investigación de la Colaboración BEME (https://www.bemecollaboration.org), y de la Colaboración Campbell (https://www.campbellcollaboration.org). En los cursos de EBE desarrollados por el grupo de la UNAM se diseñó un instrumento para evaluar investigaciones cuantitativas en educación, basado en una amplia revisión de la literatura (Cuadro 2), usando elementos de *The Research Methods Knowledge Base* (Trochim y Donnelly, 2006). Existen además recomendaciones para evaluar trabajos de corte cualitativo o mixtos, como las siguientes:
https://www.cebm.net/wp-content/uploads/2019/01/Qualitative-Studies.pdf
http://cfkr.dk/images/file/CASP%20instrumentet.pdf

Cuadro 2. Guía de análisis de trabajos de investigación cuantitativa en educación

1. Título – Resumen • ¿El título sugiere los constructos importantes y las relaciones en el estudio? • ¿El resumen proporciona información suficiente para que usted tome una decisión sobre leer el artículo completo? • ¿Se indica claramente el propósito del estudio? • ¿La introducción hace que el propósito del estudio sea fácil de comprender?
2. Introducción – revisión de la literatura • ¿Se introduce la temática del estudio en términos de lo que ya se sabe acerca del tema? • ¿Los autores describen las teorías relevantes? • ¿Señalan alguna problemática con las teorías existentes que abordarán más adelante en su estudio? • ¿Discuten hallazgos contradictorios de estudios previos? • ¿Proporcionan alguna idea en cuanto a cómo explicar esos resultados contradictorios? • ¿Se justifican las decisiones sobre el diseño y procedimientos para el estudio en función de estudios previos? • ¿Se presentan claramente los objetivos e hipótesis del estudio?
3. Método – Participantes • ¿Los temas fueron tratados bajo una perspectiva ética? • ¿El estudio fue revisado por algún comité/junta de ética? • ¿Es clara la explicación de por qué fueron seleccionados esos sujetos en particular para la muestra? • ¿Cuál fue el proceso de selección? • ¿Fue utilizada la selección aleatoria o la asignación? • ¿Se realizó un análisis de poder estadístico? • ¿Hay alguna evidencia de sesgo debido al muestreo deficiente o a la falta de respuesta (datos perdidos)? • ¿Se dan detalles de los no-participantes o desertores?

4. Método – Diseño • ¿Se menciona algún diseño de investigación en específico? • ¿Se definieron las variables principales? • ¿Se identificaron algunas amenazas particulares sobre la validez? • ¿Pasaron por alto algunas amenazas en particular de validez? • Si se trata de un estudio longitudinal ¿Justifican el número y el momento de hacer las observaciones? • Si se trata de un estudio de intervención ¿Se han descrito con claridad los procedimientos de tratamiento?
5. Método – Medidas • ¿Se referencian y describen cabalmente todos los instrumentos empleados? • ¿Se presentan datos de confiabilidad y validez de cada instrumento utilizado? • ¿Hay algún instrumento construido por el investigador del estudio? ¿Fueron piloteados? • ¿En el estudio se reporta el protocolo exacto que emplearon para la obtención de datos? • ¿Se describen los procedimientos para la obtención de puntajes? • ¿Hay algunas desviaciones de los procedimientos estándar para la obtención de puntajes?
6. Método – Análisis de datos • ¿Se describen a detalle los procedimientos para el análisis de datos? • ¿Son acorde a las preguntas que se plantearon? • ¿Se han comprobado los supuestos de las pruebas estadísticas? • ¿Hasta qué punto se han explorado los datos con respecto a irregularidades como la ausencia no aleatoria y los valores atípicos? • ¿Cómo fueron manejados los datos perdidos? • ¿Dieron referencias sobre el software estadístico que emplearon? • ¿Se describe el significado práctico o clínico? • ¿Se dan los tamaños de efecto y los intervalos de confianza? • ¿Se proporcionan los valores exactos de p? • ¿Los autores se hacen cargo de abordar toda la metodología empleada o hay ausencias?
7. Resultados y Discusión • ¿Las conclusiones del estudio se relacionan directamente con los resultados, o se extienden más allá de ellos y caen en la especulación? • ¿Las conclusiones se relacionan directamente con el propósito e hipótesis del estudio? • ¿Los resultados considerados en la discusión están vinculados al conocimiento previo? • ¿Se discuten las limitaciones del estudio?

8. Figuras • ¿El título de la figura identifica claramente las variables? ¿Los ejes de la figura están claramente etiquetados? • ¿La figura se entiende por sí misma o tiene que referirse al texto para comprenderla?
9. Tablas • ¿El título indica claramente qué variables o categorías de variables están incluidas en la tabla? • ¿Están claramente etiquetadas las filas y las columnas? ¿Se proporcionan notas con las pruebas de significancia? • ¿La tabla se entiende por sí misma o tiene que referirse al texto para entenderlo?
10. Referencias • ¿Las referencias están en el formato correcto? ¿Las referencias incluyen fuentes clave de investigadores destacados en el área?
11. Estilo de escritura • ¿Se emplea un estilo de escritura particular de manera consistente? • ¿Cuál es la calidad general de la escritura? (Considere los aspectos técnicos del estilo, así como su impresión general del documento con respecto a la calidad de la escritura.)

Adaptado de Trochim y Donnelly, 2006.

Aplicar en la práctica

El cuarto paso de la EBE es la utilización en la práctica de la información evaluada críticamente, para la solución del problema educativo que generó la pregunta o en situaciones similares subsecuentes. Es crucial cerrar el círculo virtuoso de la EBE, y darle sentido al proceso de búsqueda y análisis de la información científica publicada, para que no se convierta en un ejercicio intelectual estéril sin utilidad práctica. La aplicación del uso de la evidencia publicada tiene diferentes aproximaciones, desde el uso por el docente individual en su práctica cotidiana al dar clase en la universidad, hasta el uso institucional en políticas universitarias o gubernamentales (Nelson y O'Beirne, 2014; Sharples, 2013). Para disminuir la brecha temporal existente entre el momento de generación de la evidencia y su publicación, hasta su uso generalizado en la práctica, han aparecido nuevas modalidades de transferencia y movilización del conocimiento, como la investigación traslacional en educación (Hernández-Carrillo y Sánchez-Mendiola, 2018) y las ciencias de la implementación (Soicher et al., 2020). La aplicación de la evidencia implica su uso en un ecosistema educativo, en el que existen generadores de evidencia, sintetizadores de evidencia, distribuidores y transformadores de evidencia y, por último, implementadores de la

misma, para cerrar el círculo regresando a los generadores de evidencia (Sharples, 2013) (Figura 4). El profesorado busca diversos tipos de información para su aplicación en la práctica, y de esta manera navegar en el territorio del desarrollo profesional.

Figura 4. Elementos de un ecosistema de evidencia educativa

Productores de evidencias
Universidades
Gobiernos
Instituciones de investigación
Sector privado

Sintetizadores de evidencias
Universidades
Gobiernos
ONG / Instituciones de beneficiencia
Sector privado
Institutos de investigación

Distribuidores / transformadores de evidencias
Cuerpos profesionales
Organizaciones normativas
ONG
Medios de comunicación
Sector privado
Gobiernos locales

Responsables de la implementación de evidencias
Proveedores de servicios
Organismos / redes profesionales
Responsables locales

Adaptado de Sharples, 2013.

Evaluar la realización del proceso

El quinto paso es la evaluación de todo el proceso, desde la utilidad concreta de la información para el problema educativo específico, hasta una evaluación personal longitudinal de cómo se ejerce la docencia. De esta manera se puede mejorar el desempeño y la calidad de la docencia, de acuerdo a las necesidades generadas en la práctica personal.

CONCLUSIONES

La comunidad docente en general no está familiarizada con la existencia del material producto de investigación en educación, y tiene dificultades para encontrarlo, recuperarlo, analizarlo y utilizarlo. Es factible desarrollar habilidades y destrezas para utilizar la literatura educativa en nuestro quehacer docente cotidiano. Las universidades y centros profesionales deben adquirir el material profesional básico de educación (libros, revistas, bases de datos), entrenar a sus docentes en la búsqueda de bases de datos sobre educación y en

la evaluación crítica de artículos de las ciencias sociales, así como promover colaboración interdisciplinaria con académicos de la educación y ciencias sociales. El enfoque de la EBE para el desarrollo profesional de la práctica educativa debería integrarse en las iniciativas de mejora de calidad de universidades e instituciones académicas. En un escenario ideal, la excelencia en educación es tan importante como la excelencia en la práctica profesional y en investigación.

Existe gran cantidad de publicaciones de investigación en educación, relevante para la práctica docente. La iniciativa de EBE es de relevancia para la comunidad académica, por lo que es importante conocer este modelo para beneficio de la práctica docente y el aprendizaje de los estudiantes.

REFERENCIAS

Benavides Lara, M. A., de Agüero Servín, M., Pompa Mansilla, M., y Sánchez Mendiola, M. (2021) El curso en Educación Basada en Evidencias (EBE): reflexiones para la transdisciplina, la docencia y la investigación *DIDAC*, (78 JUL-DIC), 8-19 https://doi.org/10.48102/didac.2021..78_JUL-DIC.73

Camilli Trujillo, C., Arroyo Resino, D., Asensio Muñoz, I. y Mateos-Gordo, P. (2020). Hacia la educación basada en la evidencia: un método y un tema. *Revista Electrónica en Educación y Pedagogía, 4*(6), 69-85. Recuperado de https://doi.org/10.15658/10.15658/rev.electron.educ.pedagog20.05040606

Campillo Labrandero M, Pompa Mansilla M, Hernández Carrillo F, Sánchez Mendiola M. 2019. Una experiencia de formación: la Educación basada en evidencias. *Revista Digital Universitaria (RDU)*. Vol. 20, núm. 6 noviembre-diciembre. DOI:http://doi.org/10.22201/codeic.16076079e.2019.v20n6.a6.

Centre for Educational Research and Innovation. (2007). *Evidence in education: Linking research and policy*. Paris, France: Organisation for Economic Co-operation and Development.

Centra, J.A. (1983). Research productivity and teaching effectiveness. *Res High Educ* 18, 379–389. https://doi.org/10.1007/BF00974804

Chalmers, I. (2005) If evidence–informed policy works in practice, does it matter if it doesn't work in theory? *Evidence and Policy*. *1*(2): 227–42. https://doi.org/10.1332/1744264053730806

Cilliers, F.J., Schuwirth, L.W., Adendorff, H.J. et al., (2010). The mechanism of impact of summative assessment on medical students' learning. *Adv in Health Sci Educ* 15, 695–715. https://doi.org/10.1007/s10459-010-9232-9

Cook, D. A., & Reed, D. A. (2015). Appraising the quality of medical education research methods: the Medical Education Research Study Quality Instrument and the Newcastle-Ottawa Scale-Education. *Academic medicine: Journal of the Association of American Medical Colleges*, *90*(8), 1067–1076. https://doi.org/10.1097/ACM.0000000000000786

Dauphinee WD. & Wood-Dauphinee S. (2004). The need for evidence in medical education: the development of best evidence medical education as an opportunity to inform, guide, and sustain medical education research. *Academic Medicine: Journal of the Association of Ame-*

rican Medical Colleges 79, 925-930. https://journals.lww.com/academicmedicine/Fulltext/2004/10000/The_Need_for_Evidence_in_Medical_Education__The.5.aspx

Delandshere, G. (2001). Implicit theories, unexamined assumptions and the status quo of educational assessment. *Assessment in Education: Principles, Policy and Practice, 8*(2), 113–133. https://doi.org/10.1080/09695940123828

Diery, A., Vogel, F., Knogler, M., & Seidel, T. (2020). Evidence-Based Practice in Higher Education: Teacher Educators' Attitudes, Challenges, and Uses. *Frontiers in Education*. https://www.frontiersin.org/articles/10.3389/feduc.2020.00062/full

Duraiappah, A., van Atteveldt, N., Asah, S., Borst, G., Bugden, S., Buil, J. M., Ergas, O., Fraser, S., Mercier, J., Restrepo Mesa, J. F., Mizala, A., Mochizuki, Y., Okano, K., Piech, C., Pugh, K., Ramaswamy, R., Chatterjee Singh, N., & Vickers, E. (2021). The International Science and Evidence-based Education Assessment. *NPJ science of learning*, *6*(1), 7. https://doi.org/10.1038/s41539-021-00085-9

Education Group for Guidelines on Evaluation (1999). Guidelines for evaluating papers on educational interventions. *BMJ (Clinical research ed.)*, *318*(7193), 1265–1267. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1115651/

Eriksen, M. B., & Frandsen, T. F. (2018). The impact of patient, intervention, comparison, outcome (PICO) as a search strategy tool on literature search quality: a systematic review. *Journal of the Medical Library Association: JMLA*, *106*(4), 420–431. https://doi.org/10.5195/jmla.2018.345

Goldacre, B. (2013). *Building Evidence into Education* [online]. Disponible en https://dera.ioe.ac.uk/17530/7/ben%20goldacre%20paper_Redacted.pdf

Gorard, S., See, B.H. and Siddiqui, N. (2020), What is the evidence on the best way to get evidence into use in education? *Rev Educ,* 8: 570-610. https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1002/rev3.3200

Gough, D. (2021). Appraising Evidence Claims. *Review of Research in Education*, *45*(1), 1–26. https://doi.org/10.3102/0091732X20985072

Guerra Ferrada, C. ., y Carrasco , P. (2021). Educación médica y de ciencias de la salud basada en evidencia: perspectivas y desafíos: Evidence Based Medical Education and Health Sciences: Perspectives and Challenges. *ARS MEDICA Revista De Ciencias Médicas*, *46*(4), 66–70. https://doi.org/10.11565/arsmed.v46i4.1848

Haig, A., & Dozier, M. (2003). BEME Guide no 3: systematic searching for evidence in medical education--Part 1: Sources of information. *Medical Teacher*, *25*(4), 352–363. https://doi.org/10.1080/0142159031000136815

Haig, A., & Dozier, M. (2003). BEME guide no. 3: systematic searching for evidence in medical education--part 2: constructing searches. *Medical Teacher*, *25*(5), 463–484. https://doi.org/10.1080/01421590310001608667

Hammersley, M. (1997). Educational research and teaching. A response to Hargreaves' TTA lecture. *Br Educ Res J*. *23*(2):141-161. https://doi.org/10.1080/0141192970230203

Harden, R.M., Grant, J., Buckley, G., Hart, I.R. (1999). BEME Guide No. 1: Best Evidence Medical Education. *Med Teach*. *21*(6):553-62. https://doi.org/10.1080/01421599978960

Harden, R.M. (2000). Best evidence medical education: the simple truth. *Med Teach*. *22*(2):117-119. https://doi.org/10.1080/01421590078517

Hargreaves, D. H. (1996). Teaching as a Research-Based Profession: Possibilities and Prospects. The Teacher Training Agency, Annual Lecture. https://eppi.ioe.ac.uk/cms/Portals/0/PDF%20reviews%20and%20summaries/TTA%20Hargreaves%20lecture.pdf

Hargreaves, D. H. (1997). In Defence of Research for Evidence-Based Teaching: A Rejoinder to Martyn Hammersley. *British Educational Research Journal*, *23*(4), 405–419. http://www.jstor.org/stable/1502078

Harrison, C. J., Konings, K. D., Schuwirth, L. W. T., Wass, V., & Van der Vleuten, C. P. M. (2017). Changing the culture of assessment: The dominance of the summative assessment paradigm. *BMC Medical Education, 17*(1), 73. https://doi.org/10.1186/s12909-017-0912-5

Hart, I.R. (2000). Best evidence medical education (BEME): a plan for action. *Med Teach*. *22*(2):131-135. https://doi.org/10.1080/01421590078535

Hattie, J., & Marsh, H. W. (1996). The Relationship Between Research and Teaching: A Meta-Analysis. *Review of Educational Research*, *66*(4), 507–542. https://doi.org/10.3102/00346543066004507

Hederich Martínez, C., Martínez Bernal, J. y Rincón Camacho, L. (2014). Hacia una educación basaada en la evidencia. *Revista Colombia de Educación*, 66, 19-54. Recuperado de http://www.scielo.org.co/pdf/rcde/n66/n66a02.pdf

Hernández Carrillo, F.B., Sánchez Mendiola, M. (2018). Investigación traslacional en educación: un puente entre teoría y práctica educativa. *Revista Digital Universitaria (RDU)*. Vol. 19, núm. 4 julio-agosto. DOI: http://doi.org/10.22201/codeic.16076079e.2018.v19n4.a4

Hughes, M. (2005). The Mythology of Research and Teaching Relationships in Universities. En R. Barnett (Ed.), *Reshaping the University: New Relationships between Research, Scholarship and Teaching* (pp. 14-26). Nueva York: McGraw-Hill.

La Velle, L. (2015). Translational Research and Knowledge Mobilization in Teacher Education: Towards a "Clinical", Evidence-Based Profession? *Journal of Education for Teaching*, 41(5), 460-463. Recuperado de https://doi.org/10.1080/02607476.2015.1105534

McAleavy, T. (2015). *Teaching as a research-engaged profession: problems and possibilities. Education Development Trust*: UK. https://www.educationdevelopmenttrust.com/our-research-and-insights/research/teaching-as-a-research-engaged-profession

McKnight, L. & Morgan, A. (2020) A broken paradigm? What education needs to learn from evidence-based medicine, *Journal of Education Policy*, 35:5, 648-664, DOI: 10.1080/02680939.2019.1578902

Morrison, J. M., Sullivan, F., Murray, E., & Jolly, B. (1999). Evidence-based education: development of an instrument to critically appraise reports of educational interventions. *Medical Education*, *33*(12), 890–893. https://doi.org/10.1046/j.1365-2923.1999.00479.x

Munro N. (2004). Evidence-based assessment: no more pride or prejudice. *AACN clinical issues*, *15*(4), 501–505. https://doi.org/10.1097/00044067-200410000-00002

Nelson, J. and O'Beirne, C. (2014). *Using Evidence in the Classroom: What Works and Why?* Slough: NFER. https://www.nfer.ac.uk/publications/impa01/impa01.pdf

Newton, P. M., Da Silva, A., Berry, S. (2020). The Case for Pragmatic Evidence-Based Higher Education: A Useful Way Forward? *Frontiers in Education*, VOL 5 https://www.frontiersin.org/article/10.3389/feduc.2020.583157

Pérez Puente, Leticia. (2013). Internet para investigadores en humanidades: La búsqueda de artículos y libros en la Red. *Perfiles Educativos*, *35*(139), 190-201. http://www.scielo.org.mx/pdf/peredu/v35n139/v35n139a12.pdf

Petersen, S. (1999). Time for evidence based medical education. *BMJ (Clinical research ed.*), 318(7193):1223-4. https://doi.org/10.1136/bmj.318.7193.1223

Pigott, T. D., Tocci, C., Ryan, A. M., & Galliher, A. (2021). Quality of Research Evidence in Education: How Do We Know? *Review of Research in Education*, *45*(1), vii–xii. https://doi.org/10.3102/0091732X211001824

Procter, R. (2015). Teachers and School Research Practices: The Gaps between the Values and Practices of Teachers. *Journal of Education for Teaching, 41*(5), 464-477. Recuperado de https://doi.org/10.1080/02607476.2015.1105535

Roberts, C., Khanna, P., Lane, A. S., Reimann, P., & Schuwirth, L. (2021). Exploring complexities in the reform of assessment practice: a critical realist perspective. *Advances in health sciences education: theory and practice*, *26*(5), 1641–1657. https://doi.org/10.1007/s10459-021-10065-8

Rodriguez, M.C. (2005). Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research. *Educational Measurement: Issues and Practice, 24*(2):3-13. https://doi.org/10.1111/j.1745-3992.2005.00006.x

Sánchez-Meca, J. y Botella, J. (2010). Revisiones sistemáticas y meta-análisis: herramientas para la práctica profesional. *Papeles del Psicólogo*, *31*(1): 7-17. https://doi.org/10.1111/j.1745-3992.2005.00006.x

Sánchez-Mendiola, M. (2012). Educación médica basada en evidencias: ¿Ser o no ser? *Investigación en Educación Médica, 1*(2), 82-89. Recuperado de https://www.elsevier.es/es-revista-investigacion-educacion-medica-343-pdf-X2007505712427039

Sharples, J. (2013). *Evidence for the Frontline*. London: Alliance for Useful Evidence [online]. https://apo.org.au/sites/default/files/resource-files/2013-06/apo-nid34800.pdf

Simpson, A. (2019). The evidential basis of "evidence-based education": an introduction to the special issue, *Educational Research and Evaluation*, 25:1-2, 1-6, DOI: 10.1080/13803611.2019.1617979

Soicher, R.N., Becker-Blease, K.A. & Bostwick, K.C.P. (2020). Adapting implementation science for higher education research: the systematic study of implementing evidence-based practices in college classrooms. *Cogn. Research* **5,** 54s. https://doi.org/10.1186/s41235-020-00255-0

Trochim, W.M. and Donnelly, J.P. (2006). The Research Methods Knowledge Base. 3rd Edition, Atomic Dog, Cincinnati, OH. https://conjointly.com/kb/

Van der Vleuten, C. (1995). Evidence-based education? *Adv Physiol Educ*. *269*(6 Pt 3): S3. https://doi.org/10.1152/advances.1995.269.6.S3

Van der Vleuten, C., Dolmans, D. & Scherpbier, A. (2000). The need for evidence in education. *Medical Teacher*, 22, 246-250. https://doi.org/10.1152/advances.1995.269.6.S3

Zisblatt, L., Chen, F., Dillman, D., DiLorenzo, A. N., MacEachern, M. P., Miller Juve, A., Peoples, E. E., & Grantham, A. E. (2019). Critical Appraisal of Anesthesiology Educational Research for 2017. *Cureus*, *11*(6), e4838. https://doi.org/10.7759/cureus.4838

Sección V.

INVESTIGACIÓN E INNOVACIÓN EN EVALUACIÓN

Capítulo 42

INNOVACIONES EN EVALUACIÓN EDUCATIVA

Katherina Edith Gallardo Córdova

"La educación no consiste en llenar un cántaro sino en encender un fuego."
William Butler Yeats

INTRODUCCIÓN

Hacer un alto en el camino para comprender los tropiezos, valorar los logros y vislumbrar los retos venideros es, sin duda, un ejercicio de gran valor frente a los desafíos que se presentan en nuestros tiempos. En nuestra disciplina, pensar desde la literatura e investigaciones para transformar el proceso educativo permite, por un lado, fortalecer las buenas prácticas, así como entender las que no fueron tan favorables. Por otro lado, hace posible proponer modificaciones, realizar mejores reflexiones y actuar en aras de conseguir paulatinamente un mejor nivel de calidad educativa. Además, abre una visión prospectiva que genera expectativas sobre el futuro cercano.

Así, este capítulo, visto como un breve alto en el camino, inicia explicando lo que la sociedad de la información, la del conocimiento y ahora la sociedad del aprendizaje han ido dejando a su paso (Sánchez, 2016): hoy en día, las naciones intentan avanzar a gran velocidad en la tarea de potenciar al máximo el desarrollo del talento humano, para lo cual utilizan los avances tecnológicos para flexibilizar el proceso enseñanza-aprendizaje, en diferentes contextos y con distintas intenciones formativas. Es, por tanto, importante vislumbrar que ese es el reto actual de las naciones: poner en marcha nuevos procesos educativos que respondan a las demandas formativas de los ciudadanos del siglo XXI, en un mundo que experimenta, de forma constante, cambios acelerados.

Lo anterior, nos hace pensar en la evaluación del aprendizaje como uno de los eslabones del proceso formativo que permite tomar decisiones a los docentes para enfrentar estas transformaciones mundiales en la educación. Es preciso destacar que no solo la instrucción ha experimentado la inserción de nuevas ideas, estrategias o tecnologías. Sin duda, la evaluación es uno de los procesos que mayor atención ha cobrado en los últimos años, vinculado con las nuevas tenencias en innovación educativa y el uso de TIC para su diseño, aplicación y análisis (Capllonch Bujosa y Buscà Donet, 2017; Elizondo-Garcia et al., 2019; Hudha et al., 2018).

Es preciso considerar que cada día contamos con mayor información que emana de investigaciones educativas y tecnológicas que permean en la creación de nuevas alternativas para conducir procesos de evaluación. Sin embargo, hay muchas preguntas que emergen en este actual punto de la historia, las cuales requieren respuestas para seguir avanzando. De hecho, el evento mundial que más preguntas suscitó sobre la evaluación de aprendizaje ha sido el paso de la pandemia por COVID-19 y la emergencia educativa que se suscitó a raíz de este problema global (CEPAL-UNESCO, 2020; Martín-Sómer et al., 2021; Miguel, 2020).

Así, la situación de emergencia educativa ha traído consigo reflexiones importantes sobre las competencias docentes necesarias para afrontar los retos del cambio, pasando de una modalidad totalmente presencial a ambientes en línea o combinados. Las dudas relacionadas con la implementación de estrategias para la evaluación del aprendizaje giraron, principalmente, en función del cuidado de la integridad académica, el diseño de instrumentos con apoyo de aplicaciones, de la eficacia y eficiencia al momento de retroalimentar, de la priorización de actividades sin poner en riesgo el rigor académico, entre los principales cuestionamientos. Se estima un despliegue de aproximadamente 15 mil anuncios en internet sobre cursos, talleres y *webinars* en idioma español en apoyo al desarrollo de competencias docentes en formato en línea y combinados, de todos los niveles educativos. Este despliegue se ha venido realizando con apoyo de las redes sociales, un fenómeno de capacitación docente nunca antes visto *(Google Trends,* 2021*)*. De los aspectos más positivos dentro de estas fatídicas circunstancias para los sistemas escolares, se destaca el avance que se ha presentado el ejercicio de competencias digitales y las reflexiones realizadas sobre la toma de decisiones en evaluación del aprendizaje.

Por lo anterior, para realizar un análisis de lo que ha significado la innovación educativa en evaluación del aprendizaje. En este capítulo se presenta un mapeo, así como una revisión sistemática de literatura, con el fin de fotografiar de varias formas lo que ha estado sucediendo en esta temática, durante los últimos cinco años en diferentes regiones del mundo. Este análisis va acompañado de respuestas a preguntas clave, con tintes de presente y prospectiva. La finalidad es consolidar información que nos permita responder en qué punto del camino estamos, qué ha caracterizado nuestro avance y qué expectativas se tienen sobre el avance en este tema en los siguientes años.

DESARROLLO

ESTUDIO DE LA INNOVACIÓN EDUCATIVA Y LA EVALUACIÓN DEL APRENDIZAJE EN LOS ÚLTIMOS AÑOS

El tema de innovación educativa ha sostenido una tendencia de consulta al alza, en cuanto a búsquedas de información para su mejor entendimiento, manejo y experimentación. Las estadísticas sobre las tendencias, marcadas por uno de los buscadores más utilizados del mundo, permitió obtener un cálculo acera de las búsquedas realizadas entre los años 2016 y 2021 sobre “innovación educativa” y su contraparte en idioma inglés “*educational innova-*

tion" en el campo educativo. Se encontró que, en América Latina, las naciones desde donde se presentaron más búsquedas son: Ecuador, México, España, Guatemala y Perú. En cuanto a la misma búsqueda en inglés, se encontraron los países: Nigeria, Filipinas, Estados Unidos, India y Tailandia.

Este mismo ejercicio, pero con las palabras "evaluación del aprendizaje" y su contraparte en idioma inglés "*learning assessment*" arrojó que los países desde donde se presentaron más búsquedas son: Panamá, Ecuador, República Dominicana, Perú y Guatemala en el caso de países hispanoparlantes. En el resto del mundo, con uso de la palabra clave en inglés, se encuentran: Filipinas, Zimbabue, Etiopía, Namibia y Sudáfrica. En ambos casos, coincide que el periodo en el que se presentan más búsquedas es entre los meses de marzo y julio del año 2020 (Google Trends, 2021).

No obstante, el ejercicio anterior puede permitirnos obtener un primer panorama mundial para estimar la diversidad de países desde donde las temáticas de innovación educativa y evaluación del aprendizaje se consultan con mayor frecuencia, es necesario pasar a ejercicios más focalizados y de corte científico, para entender un poco mejor la producción de obras y abordaje de temáticas específicas.

Por lo anterior, para poder determinar una serie de búsquedas apropiadas para ilustrar la fusión entre innovación educativa y evaluación, se presenta en la Figura 1 que engloba una serie de palabras clave que emanan de la lectura de varios estudios (Prendes y Cerdán, 2021; Ramírez y García-Peñalvo, 2018; Riccomini et al., 2019), a través de las cuales se busca explorar nuevos horizontes en el tema de evaluación del aprendizaje desde la innovación y, a su vez, discutir en función de una prospectiva.

Figura 1. Palabras clave utilizadas para las búsquedas de la literatura que derive en una visión global del presente y una prospectiva en innovación educativa y evaluación del aprendizaje



Nota. Palabras usadas: innovación educativa, evaluación del aprendizaje, aprendizaje adaptativo, inteligencia artificial, trabajo colaborativo, realidades extendidas, gamificación, aula invertida y analíticas de aprendizaje, con sus equivalentes en inglés.

Para enriquecer el análisis desde un panorama mundial y, a la vez, tener una visión más actual de las diferentes perspectivas de trabajo en el tema de innovación en educación superior, se tomó como referencia un estudio publicado por Riccomini et al. (2019) que enuncia y explica 11 diferentes prácticas desde las que las instituciones de educación superior (IES) están trabajando en innovación educativa. Aunque se indica dentro de estas 11 prácticas una específicamente dedicada a la evaluación del aprendizaje, es común encontrar en la literatura que la evaluación del aprendizaje se combina con algunas otras prácticas dentro del ámbito de la innovación. Entender esta fusión de intereses desde la literatura puede apoyar a entender con mayor claridad los esfuerzos actuales y las tendencias al corto y mediano plazo.

Figura 2. 11 prácticas desde las que las instituciones de educación superior (IES) están trabajando en procesos de innovación educativa



Riccomini et al., 2019.

VISIÓN GLOBAL

El resultado de la búsqueda de literatura en el tema de innovación educativa y evaluación del aprendizaje entre los años 2016 y 2021, permitió seleccionar 21 obras (artículos de investigación o ponencias en congresos internacionales) que, por su relación directa con la temática, la aparición de palabras clave anteriormente expuestas (Figura 1) y/o la citación con la que cuenta a la fecha, expliquen en sí mismas la esencia de las inquietudes actuales y, a la vez, exponen tendencias educativas para los años venideros. Cabe destacar que se consideró necesario que este análisis integrara publicaciones en inglés o en español, en aras de enriquecer la discusión posterior.

Así, en la Tabla 1, se encuentran el resultado final de selección a partir de la búsqueda por temáticas y el número de obras seleccionadas en diferentes bases de datos como Scopus, Web of Science y un motor de búsqueda como Google Académico. Se seleccionaron estas obras por considerarlas relevantes para entender desde la fusión de temáticas, entre innovación educativa y evaluación del aprendizaje, lo que subyace en este campo de estudio. En el caso de las obras en español, se determinó su inclusión en este análisis especialmente por la afinidad de los temas, no así por la citación alcanzada.

Tabla 1. Integración de obras que permiten realizar un análisis sobre la visión actual y prospectiva en innovación educativa y evaluación del aprendizaje

Palabras clave en inglés y español	Autor(es)	Título	Citas
Learning assessment AND educational innovation	(Keinänen et al., 2018)	How to measure students' innovation competences in higher education: Evaluation of an assessment tool in authentic learning environments.	40
	(Coates, 2016)	Assessing student learning outcomes internationally: insights and frontiers.	23
Innovación educativa AND evaluación del aprendizaje	(Martínez et al., 2017)	La innovación educativa y la evaluación en línea como fortalecimiento de la enseñanza superior.	0
Adaptive learning OR personal learning environment AND learning assessment	(S. Gupta et al., 2021)	Use of an Adaptive e-Learning Platform as a Formative Assessment Tool in the Cardiovascular System Course Component of an MBBS Programme.	2
	(Tempelaar, 2020)	Supporting the less-adaptive student: the role of learning analytics, formative assessment and blended learning.	10
Aprendizaje adaptativo OR ambiente personalizado de aprendizaje AND evaluación del aprendizaje	(González et al., 2017)	Aprendizaje Adaptativo: Un Caso de Evaluación Personalizada.	11
Artificial intelligence AND learning assessment	(Gardner et al., 2021)	Artificial intelligence in educational assessment: Breakthrough? Or buncombe and ballyhoo?'	1
	(Sekeroglu et al., 2019)	Artificial Intelligence in Education: application in student performance evaluation.	7
Inteligencia artificial AND evaluación del aprendizaje	(Castrillón et al., 2020)	Predicción del rendimiento académico por medio de técnicas de inteligencia artificial.	24
Collaborative work AND learning assessment	(Sridharan et al., 2019)	Does the use of summative peer assessment in collaborative group work inhibit good judgement?	47
	(Heitkamp et al., 2018)	Work shadowing in dental teaching practices: evaluation results of a collaborative study between university and general dental practices.	32
Aprendizaje colaborativo AND evaluación del aprendizaje	(Molina et al., 2016)	Innovación docente en Educación Superior: edublogs, evaluación formativa y aprendizaje colaborativo.	37

Extended reality AND [virtual reality, augmented reality] learning assessment	(Gupta et al., 2019)	A Virtual Reality Enhanced Cyber- Human Framework for Orthopedic Surgical Training.	21
	(Chu et al., 2019)	Effects of formative assessment in an augmented reality approach to conducting ubiquitous learning activities for architecture courses.	40
Realidades extendidas [realidad virtual, realidad aumentada] AND evaluación del aprendizaje	(Cabero Almenara et al., 2017) (Cabero Almenara et al., 2019)	Dispositivos móviles y realidad aumentada en el aprendizaje del alumnado universitario.	105
Gamification OR serious game AND learning assessment	(Díaz- Ramírez, 2020)	Gamification in engineering education – An empirical assessment on learning and game performance.	7
	(Handoko et al., 2021)	Gamification in Learning using Quizizz Application as Assessment Tools.	1
Gamificación OR juego en serio AND evaluación del aprendizaje	(Rodríguez et al., 2019)	Metodología innovadora basada en la gamificación educativa: evaluación tipo test con la herramienta Quizizz.	22
Flipped classroom AND learning assessment	(Morton y Colbert-Getz, 2017)	Measuring the impact of the flipped anatomy classroom: The importance of categorizing an assessment by Bloom's taxonomy.	53
	(Ojennus, 2016)	Assessment of learning gains in a flipped biochemistry classroom.	77
Aula invertida AND evaluación del aprendizaje	(Maluenda Albornoz et al., 2021)	Efectos del aula invertida y la evaluación auténtica en el aprendizaje de la matemática universitaria en estudiantes de primer año de ingeniería.	2

ACTUALIDAD DESDE SIETE TEMÁTICAS CONTEMPORÁNEAS QUE REÚNEN A LA INNOVACIÓN EDUCATIVA Y A LA EVALUACIÓN DEL APRENDIZAJE

A partir de la selección de literatura, se distinguen siete temáticas que permitirán exponer el panorama actual de la innovación educativa en su relación estrecha con la evaluación del aprendizaje. A continuación, se desarrollan una serie de ideas clave por cada temática para presentar tanto el panorama actual como el prospectivo desde las obras seleccionadas.

Tema 1: Innovación educativa en los procesos de evaluación del aprendizaje

Luego de revisar a Keinänen et al. (2018), Coates (2016) y a Martínez et al. (2017), en este primer tema se pueden identificar ciertos factores comunes alrededor de la discusión en el tema de innovación en evaluación del aprendizaje. El primero se relaciona con una preocu-

pación latente sobre cómo, desde las IES, se mide el desarrollo de competencias que exige la sociedad del conocimiento y el aprendizaje. Las prácticas de evaluación, concretamente realizadas a través de la valoración de evidencias de aprendizaje o resultados de las examinaciones, no están teniendo el mismo nivel de transformación que sí se está presentado en los procesos de diseño curricular e instruccional. Esta situación pone en riesgo la congruencia en la presentación de resultados y, por ende, en la rendición de cuentas sobre las capacidades con la que los egresados se insertan en el mundo laboral, donde se espera un desempeño acorde con las demandas sociales y económicas actuales.

Un segundo factor se relaciona con la apuesta a la autonomía y la reflexión como elementos que coadyuvan a innovar los procesos de evaluación. Así, la autoevaluación se rescata en estos últimos años como una práctica de alto valor que coadyuva a la autorregulación y a fortalecer las estrategias metacognitivas, lo cual implica abandonar las creencias de ejercer la autoevaluación solo como un mecanismo de auto calificación.

Un tercer factor se relaciona con la necesidad de fortalecer el capital humano y tecnológico en las IES para el ejercicio de la evaluación. Para ser eficaces y efectivos, se apela al trabajo en múltiples niveles, que implique involucrar a los docentes, a los gestores y líderes institucionales, así como a contrapartes externas, que resulten relevantes para el desarrollo estudiantil. Este fortalecimiento se enfoca a entender lo que implica el desarrollo de las competencias docentes para evaluar (*assessment literacy*, por su denominación en inglés) y cómo enfocarse en estas competencias impactaría en el avance de los niveles de innovación educativa anhelados.

Estos tres factores, vistos desde las 11 prácticas de innovación (Riccomini et al., 2019), ubican estos esfuerzos de innovación, no solo en los métodos de evaluación, sino en el soporte a los estudiantes, en el uso de TIC, en la administración organizacional y en los métodos de enseñanza, entre las principales.

Tema 2: Aprendizaje adaptativo, aprendizaje personalizado y la evaluación del aprendizaje

Luego de analizar las obras de Gupta et al. (2021), Tempelaar (2020) y González et al. (2017), es indudable que el tema de aprendizaje adaptativo se ha convertido en un foco de atención cada vez más recurrente en IES, como alternativa para flexibilizar y personalizar el aprendizaje en sistemas educativos que buscan la igualdad de oportunidades, el estricto control del rigor académico y, en paralelo, procesos cada vez más transparentes en cuanto a rendición de cuentas. El aprendizaje adaptativo utiliza técnicas para ejercer la adaptabilidad. En primer lugar, el educador diseña las instrucciones para que el alumno pueda dominar paulatinamente el contenido. En segundo, el docente se apoya en la adaptabilidad algorítmica que coadyuva a identificar el grado de conocimiento del alumno y, por tanto, administra la mejor secuencia de instrucciones y estrategias de formación para proporcionar una experiencia de aprendizaje personalizada que le guíe hacia el dominio del contenido. Por lo anterior, se entiende que la adaptación del aprendizaje apuesta por un proceso formativo centrado en el estudiante.

Actualmente, se cuenta con el apoyo de múltiples TIC diseñadas exprofeso con esta intención (ejemplo: Aleks, Coogbooks, SchootPad, Fireckacker, por mencionar algunas) muchas de ellas relacionadas con alguna disciplina y con capacidad de trabajarse desde las plataformas de gestión del aprendizaje (LMS, por sus siglas en inglés). Estas aplicaciones ofrecen a docentes y estudiantes dosificar los contenidos, así como la aplicación de mecanismos de evaluación y retroalimentación paulatinamente, en aras de mejorar el desempeño general del ambiente de aprendizaje, y en específico, del logro de las metas de aprendizaje de cada estudiante.

Otro factor importante que se ha estado trabajando dentro de esta personalización, es la comprensión de las diferentes variables que convergen y se cristalizan en detectar diferencias sustanciales en los conocimientos previos con los que llegan los estudiantes a una clase. Estos factores de variabilidad van desde los tipos de oportunidades de formación previa, la movilidad nacional o internacional que hayan tenido los estudiantes, los cambios de modelo educativo por lo que hayan transitado, entre los principales.

Nuevamente, en este tema, emerge la autoevaluación como una práctica deseable que permite a los estudiantes analizar sus niveles de desempeño y actuar hacia el logro del rendimiento esperado. Además, intervienen otros constructos que coadyuvan a la personalización, como estilos de aprendizaje, introducido como una alternativa para realizar diseños instruccionales y evaluativos más acorde con las características de los aprendices y sus preferencias para aprender.

Sin embargo, es claro que las tecnologías y algoritmos cada vez más precisos que permiten personalizar el aprendizaje y la evaluación no son elementos suficientes para la adaptación. Pensar que por sí solas las aplicaciones realizarán lo necesario para cerrar la brecha entre el desempeño logrado y el deseado, dista de la realidad. Por lo tanto, en los estudios se hace especial énfasis en la calidad de la retroalimentación oportuna del aprendizaje, con apoyo en las analísticas para mejorar el contenido de los comentarios sobre el desempeño de cada estudiante.

Tema 3: Inteligencia artificial y evaluación del aprendizaje

Las aportaciones de Gardner et al. (2021), Sekeroglu et al. (2019) y Castrillón et al. (2019), relacionadas con el uso de inteligencia artificial (IA) en la enseñanza y el aprendizaje, permiten entender que esta tecnología ha ganado visibilidad y relevancia en la última década. La gran aportación de la IA al campo educativo se centra en la generación de patrones de conducta o acción que permiten perfilar un desempeño. Por tanto, el poder predictivo de estos datos coadyuva a generar estrategias de mejora invaluables en la toma de decisiones antes y durante el proceso formativo.

Luego de la revisión de los artículos seleccionados, hay tres puntos relevantes a tratar en el tema de IA en evaluación del aprendizaje. El primero se trata de la incursión de esta tecnología en la retroalimentación. Se reportan avances en el uso de las analíticas del aprendizaje para guiar de forma personalizada a cada estudiante. Este proceso se está volviendo especialmente útil cuando se tiene la opción de analizar amplias bases de datos de aprendizaje, como, por ejemplo, en la oferta de MOOC. Es justamente en estos entornos, donde

se vuelve muy costoso dar retroalimentación persona a persona a miles de aprendices que toman estos cursos desde diferentes lugares del planeta. Es en este caso donde la IA se vuelve crucial para orientar y dar retroalimentación a los participantes. Sin embargo, como en el caso del uso de otras tecnologías, se requieren apoyos complementarios por parte de los docentes. En este caso, se recomienda generar apoyos genéricos (blogs, videos, materiales auxiliares) en función de los errores comunes que estos sistemas detectan para afianzar el proceso de aprendizaje.

Una segunda aportación de las tecnologías basadas en IA a la evaluación del aprendizaje se encuentra también en la retroalimentación, pero de procesos específicos como puede ser la escritura de textos. La IA, actualmente, integra patrones que se presentan en la comunicación escrita, lo cual ha permitido emitir una serie de criterios con base en los cuales los estudiantes pueden recibir una retroalimentación a sus textos escritos, que van desde elementos escriturales (vocabulario, sintaxis, elementos gramaticales, formalidad, uso del idioma) hasta aspectos emocionales (posibles reacciones del lector, sentimientos transmitidos, etcétera.).

Una tercera aportación va más allá del producto generado, sean respuestas a exámenes, elaboración de ensayos o actividades de aprendizaje. Esta tercera aportación de la IA se centra en el registro de marcas de tiempo, movimientos y expresiones que se generan cada vez que nuestros estudiantes desarrollan actividades de aprendizaje. Este tipo de registros podría inclusive indicar cómo resuelven un problema, los errores cometidos y los que potencialmente podría cometer, la aplicación errónea de conceptos e incluso su capacidad de recuperación ante la frustración del poco o nulo avance en su aprendizaje. La amplia gama de dispositivos móviles con los que se cuenta hoy en día (celulares, pulseras relojes inteligentes, etcétera.) son potenciales registradores de movimientos oculares, expresiones faciales, postura corporal, gestos y discurso que se presentan en estas situaciones.

Tema 4: Aprendizaje colaborativo y evaluación del aprendizaje entre pares

Una de las competencias que la sociedad del conocimiento y el aprendizaje demanda actualmente es la capacidad de trabajar colaborativamente. Como se indica en las obras de Sridharan et al. (2019), Heitkamp et al. (2018), así como en la de Molina et al. (2016), en diferentes continentes, como Asia, Oceanía y Europa, el trabajo colaborativo es motivo de diálogo frecuente al tratar el papel de la educación superior como responsable del desarrollo de esta competencia. Se sabe que los desafíos laborales que enfrentan los egresados de carreras profesionales son cada vez más complejos en cuanto a la comunicación y la toma de decisiones en equipo, por tanto, la preocupación por el desarrollo de esta competencia permea desde la fuerza laboral hacia las IES.

Dentro de las actividades formativas que se realizan para el desarrollo del trabajo colaborativo, se han investigado innovaciones en torno a la aplicación de la evaluación del desempeño entre pares. Esta práctica se está presentando tanto en modalidades presenciales como en línea o combinadas. Aunque la práctica permite a los aprendices, por un lado, analizar los desempeños de otros y, a su vez, contar con juicios de sus pares como parte del proceso de aprendizaje, un primer punto de estudio a destacar de este proceso es que se desconoce en

gran medida la precisión y consistencia del proceso de calificación de los compañeros, especialmente cuando se les brinda el poder de premiar o penalizar a sus pares durante los regímenes de evaluación formativa y sumativa. Un hallazgo en los estudios reportados es que, si bien es cierto, los estudiantes son capaces de juzgar con precisión y coherencia el rendimiento de sus compañeros en la gran mayoría de los casos, los resultados también demuestran la existencia de un sesgo importante en la calificación de los compañeros cuando las notas de éstos contribuyen a la calificación final.

Además de los procesos relacionados con la evaluación y retroalimentación entre pares, en los últimos años también han emergido tecnologías que coadyuvan a realizar evaluaciones alrededor del desarrollo de esta competencia. Uno de los artículos reporta el uso de SPARKPLUS. Este es un sistema de autoevaluación y de evaluación entre pares que ofrece la oportunidad de calificarse a sí mismos y a sus compañeros de equipo. Este proceso de calificación arroja información sobre el estudiante y sus pares pueden mejorar su capacidad para trabajar colaborativamente. Para ello es necesario desarrollar e integrar en el sistema una lista de criterios que midan el trabajo colaborativo.

Cabe destacar que la tendencia apunta hacia el desarrollo de los criterios. El trabajo colaborativo demanda el ejercicio y regulación de una serie de recursos personales y conocimiento, entre los que destacan: la personalidad, el carácter, la comunicación, el dominio disciplinar, el liderazgo, la orientación a la tarea, la mediación, por mencionar algunos.

Tema 5: Realidades extendidas y evaluación del aprendizaje

Conforme lo indican las investigaciones de Gupta et al. (2019), Chu et al. (2019) y Cabero Almenara et al. (2019) con mayor frecuencia, encontramos a nuestro paso situaciones de aprendizaje en las que se utilizan una serie de recursos digitales que permiten recrear entornos u objetos en formato digital, con alta similitud a las características que se presentan en ambientes reales. Esta tecnología denominada realidades extendidas (realidad virtual, realidad aumentada, video 360°, entre las más populares) permea en las aulas universitarias. El acceso, cada vez menos costoso, al uso de gafas especiales, acceso a plataformas especiales para interactuar en entornos virtuales y aplicaciones más sofisticadas para dispositivos móviles, aceleran el uso de estas tecnologías en el campo educativo.

Una de las realidades extendidas que se está usando más frecuentemente en las aulas es la realidad aumentada, considerada de utilidad para guiar a los estudiantes a desempeñarse paulatinamente en el mundo real con apoyo de los recursos de simulación. En el caso específico del estudio de Chu et al. (2017) se encontró relación positiva entre el uso de esta tecnología y las prácticas de evaluación formativa en estudiantes de arquitectura. Esto permitió considerar, desde los resultados de un estudio cuasi-experimental, que se pueden obtener mejores resultados en el rendimiento de estudiantes que utilizaron este tipo de apoyos para su aprendizaje.

En el estudio que reportan Cabero Almenara et al. (2019) con respecto a la realidad aumentada, aunque el foco de atención aparentemente es el nivel de motivación, el estudio permite ver también las repercusiones de esta alta motivación en el rendimiento académico.

En el caso del estudio de Gupta et al. (2019) respecto al uso de la realidad virtual, estos recursos de aprendizaje con simulaciones que emulan la realidad y la interacción compleja que se da entre sus elementos, también reporta resultados positivos en términos de evaluación del aprendizaje y el desempeño, en aprendizajes que requieren práctica y precisión como en el caso de las ciencias de la salud. Es preciso indicar que tanto el enfoque formativo centrado en la información para planificar, diseñar y construir los simuladores como los modelos de formación para la ejecución de procesos quirúrgicos permitieron la ejecución de entrenamientos nunca antes logrados dentro del entorno áulico. Además, cabe destacar que los sistemas se apoyan en la guía de expertos (en este caso médicos de amplia trayectoria) que desempeñaron un papel importante en el diseño de los entornos de formación y en la conducción de actividades. En este caso, una vez más, se comprende que el proceso de evaluación no solo se enriquece con el uso de las TIC, sino con el acompañamiento y supervisión de formadores expertos en la disciplina, lo que resulta fundamental para lograr un mejor proceso de valoración y retroalimentación.

Tema 6: Gamificación, juego en serio y evaluación del aprendizaje

Se sabe que tanto la gamificación como el juego serio comparten la intención de añadir elementos de juego al proceso formativo, en aras de aumentar la motivación y las emociones, como elementos que contribuyen al logro de una meta específica de aprendizaje. De hecho, son estrategias que se están insertando fuertemente en procesos formativos en la última década. Las investigaciones de Díaz Ramírez (2020), Handoko et al. (2021) y Rodriguez et al. (2019) son ejemplos del creciente interés por entender más sobre los beneficios e implicaciones pedagógicas que recaen en la decisión de insertar este tipo de actividades lúdicas en clase, con apoyo de TIC.

La evaluación del aprendizaje es pieza clave del proceso lúdico, la cual se contempla tanto desde la planificación como en la ejecución del juego. Así, es posible afirmar que la evaluación es un elemento que apoya el éxito de la enseñanza y el aprendizaje cuando se aplican estas estrategias.

Un elemento común en las investigaciones que buscan entender esta innovación en el aula es que se intenta justificar y probar los efectos de causalidad entre la introducción de elementos de juego en la actividad de aprendizaje y, por tanto, el reflejo de logro de un mejor desempeño. En los últimos años, además de la introducción de elementos de competitividad y reconocimiento dentro de estas estrategias, también se está estudiando el factor de interacción y el relacionado con el trabajo colaborativo que se debe ejercer en dinámicas de juego en equipo. Ambos, son reconocidos como elementos que coadyuvan al éxito de las estrategias lúdicas en el aprendizaje.

Este dos de los casos seleccionados para este análisis, se encuentra reportado el uso de la herramienta *Quizizz*. Según se describe, este recurso permite que muchas personas trabajen en problemas, de forma simultánea, con apoyo de una pantalla interactiva y con una interfaz que demanda actividad catalogada como divertida. En cuanto a los resultados de usar esta aplicación, se indica una correlación positiva entre el uso del programa con la obtención

de mejores índices de aprobación en comparación con el hecho de no usar esta aplicación. Además, las recompensas que se emiten durante las contiendas (insignias, puntos, etcétera.) fueron motivadores que impulsaron la participación activa en el juego. Sin embargo, hay áreas de oportunidad respecto al proceso de evaluación, por ejemplo, la precisión de la retroalimentación del aprendizaje que se entrega luego del proceso de juego.

Tema 7: Aula invertida y evaluación del aprendizaje

La adopción del aula invertida ha cobrado relevancia en los últimos años. Esta metodología se ha distinguido en los últimos años como factible de ejecutarse en diversos escenarios educativos gracias a una serie de funciones que permiten el diseño de materiales, grabaciones y alojamiento de material audiovisual en sitios web, a los que tanto profesores como estudiantes tienen acceso a través de sus computadoras personales. En las investigaciones revisadas tanto Morton y Colbert-Getz (2017), Ojennus (2016) y Maluenda Albornoz et al. (2021) destacan las enormes ventajas que representa dejar de lado la clase expositiva, dado que los estudiantes revisaron a detalle los aspectos teóricos en casa, para aprovechar el tiempo áulico en la aplicación de estrategias de aprendizaje activo como: debates, conferencias interactivas, resolución de problemas, experimentos de laboratorio, juegos de rol, diseño y creación colaborativa, entre las principales.

Cabe destacar que los estudios revisados coinciden en que se presentan resultados positivos en el aprendizaje al integrar la metodología de aula invertida. No obstante, vale la pena destacar algunas particularidades de estos estudios que llaman la atención en cuanto la introducción de innovaciones relacionadas con el ejercicio de la evaluación del aprendizaje en estos escenarios.

En los estudios de Morton y Colbert-Getz (2017) y de Ojennus (2016) los investigadores regresan a las bases del diseño instruccional y su conexión con la evaluación del aprendizaje para diseñar las lecciones dirigidas a estudiantes de medicina. Para tal fin, integraron el uso de la taxonomía de Bloom para analizar las diferencias en el rendimiento cuando solicitaban a los estudiantes transitar por diferentes niveles de pensamiento luego de la exposición al uso de aula invertida para el estudio. Son interesantes las aportaciones de ambos estudios, en tanto se concluye que no basta con determinar buenos materiales de revisión independiente y luego realizar evaluaciones. Hay que integrar el factor nivel de dificultad en función de los procesos de pensamiento para tomar mejores decisiones sobre lo que habría que trabajar y reforzar en las horas de interacción en el aula.

En el estudio de Maluenda Albornoz et al. (2021) se destaca la integración del concepto de evaluación auténtica, sumado a los factores de revisión previa y aprendizaje activo durante la actividad áulica. Esta evaluación auténtica se define, en este caso, como aquella que permite conectar teoría y práctica, desechando la idea de que la evaluación del conocimiento en el aula invertida debería ejercerse desde el planteamiento de preguntas de memorización. Por lo anterior, los autores diseñaron un proceso de resolución de problemas guiado, lo que permitiría ejercer procesos de evaluación auténtica. Se reportan resultados positivos, sobre el rendimiento alcanzado, así como la satisfacción de estudiantes de educación superior alrededor de la metodología de enseñanza y evaluación.

PROSPECTIVA

Luego de haber realizado un breve análisis desde estudios publicados en inglés y español, que relacionan procesos de innovación educativa con la evaluación del aprendizaje en los últimos cinco años, es posible entrever algunas tendencias que marcarán estas investigaciones en un futuro cercano. Se juzga importante iniciar con una corta reflexión, trayendo nuevamente la frase de William Butler Yeats: *"La educación no consiste en llenar un cántaro sino en encender un fuego"*. Se vislumbra en los estudios revisados que los procesos de innovación en la evaluación del aprendizaje apuntan a continuar comprendiendo cómo encender el fuego del aprendizaje. Este fuego seguirá avivándose desde la curiosidad intelectual de docentes e investigadores, impulsados por la necesidad de conocer más sobre papel de la motivación y las emociones como elementos presentes en las vivencias de aprendizaje. Estos estudios incluirán, sin duda, algunas aproximaciones que funjan como catalizadores para el logro de un mejor rendimiento y factores que coadyuvan a alcanzar altos niveles de desempeño.

Otra tendencia que se puede vislumbrar es la personalización del aprendizaje y, por tanto, de la evaluación. Se requieren cambios de estrategia en las decisiones de diseño instruccional y evaluación, además de más indagación sobre la estructuración de este binomio. El estudio de la personalización de los ambientes de aprendizaje y la evaluación se enriquecerá mucho más en los siguientes años con el avance de la IA, que permitirá seguir generando patrones de conducta más detallados y, por lo tanto, ser capaces de postular estrategias que permitan actuar oportunamente ante el posible fracaso académico. Esto traerá avances en el diseño de aplicaciones cada vez más amigables y accesibles a docentes de diferentes partes del mundo.

Una tercera tendencia que emerge de este análisis se vincula con el papel de la interacción y el trabajo colaborativo, tanto en la innovación educativa como en la evaluación del aprendizaje. El aprendizaje, como el resultado de la construcción del conocimiento con pares, aprender del otro y reflexionar de manera conjunta ante el reto de dar solución a tareas complejas, son sin duda aspectos que serán motivo de más estudios de evaluación en diferentes contextos formativos. Aparentemente, en ambientes presenciales o metaversos, la evaluación del desempeño desde las competencias inmersas en el trabajo colaborativo serán foco de atención en los años venideros.

CONCLUSIONES

La innovación educativa en evaluación del aprendizaje se ha convertido en un campo de estudio cada vez más amplio. Puede afirmarse que es un área de creciente interés para la comunidad científica. Desde las 11 prácticas de educación superior (IES) conforme lo presentaron Riccomini et al. (2019), se vislumbran varias aristas de la innovación en donde amerita incluir estudios y avances en evaluación del aprendizaje. Así, desde hace varios años, el interés de la fusión de estas temáticas ha ido avanzando de forma sostenida, yendo más allá del estudio centrado en el rendimiento académico como fin último de la evaluación. De hecho, el foco de atención ha ido rotando su eje de cara a enfocar el trabajo de investigación

alrededor de la comprensión más profunda sobre el desarrollo de competencias disciplinares y transversales o genéricas, y por tanto, con miras a entender más sobre la evaluación de desempeño y auténtica, las estrategias de enseñanza y los cambios curriculares necesarios para lograr concordancia entre la instrucción y la evaluación.

Además, se anticipa que la comprensión de patrones de conductas y preferencias, se presenta como una estrategia prometedora para la mejora de la vida del estudiante, hacia el logro de aprendizajes que perduren en el tiempo. Esto permite afirmar que el papel de la IA será crucial en los siguientes años, tanto para la investigación aplicada como para el desarrollo de soluciones educativas.

Por todo lo anterior, la invitación a la comunidad educativa (docentes, directivos, organizaciones vinculadas a las IES) es continuar soñando, trabajando e investigando en innovación y evaluación del aprendizaje para avivar el interés por aprender. Hay que tener presente que, con nuestras decisiones pedagógicas, podemos hacer que la experiencia de aprendizaje de nuestros estudiantes sea memorable y trascienda en su vida profesional y personal.

REFERENCIAS

Cabero Almenara, J., Fernández Robles, B., y Marín Díaz, V. (2017). Dispositivos móviles y realidad aumentada en el aprendizaje del alumnado universitario. *RIED. Revista Iberoamericana de Educación a Distancia*, *20*(2), 167. https://doi.org/10.5944/ried.20.2.17245

Cabero Almenara, J., Martínez Pérez, S., Valencia Ortiz, R., Leiva Nuñez, J. P., Orellana Hernández, M. L., y Harvey López, I. (2019). La adicción de los estudiantes a las redes sociales on-line: un estudio en el contexto latinoamericano. *Revista Complutense de Educación*, *31*(1), 1–12. https://doi.org/10.5209/rced.61722

Capllonch Bujosa, M., y Buscà Donet, F. (2017). La evaluación formativa como innovación. Experiencias en una facultad de formación del profesorado. *Psychology, Society, & Education*, *4*(1), 45. https://www.researchgate.net/publication/318746633

Castrillón, O. D., Sarache, W., y Ruiz-Herrera, S. (2020). Predicción del rendimiento académico por medio de técnicas de inteligencia artificial. *Formación Universitaria*, *13*(1), 93–102. https://doi.org/10.4067/s0718-50062020000100093

CEPAL-UNESCO. (2020). Educación en tiempos de pandemia (COVID-19). *Revista Universidad de La Salle*, *1*(85), 51–59. https://doi.org/10.19052/ruls.vol1.iss85.4

Chu, H. C., Chen, J. M., Hwang, G. J., & Chen, T. W. (2019). Effects of formative assessment in an augmented reality approach to conducting ubiquitous learning activities for architecture courses. *Universal Access in the Information Society*, *18*(2), 221–230. https://doi.org/10.1007/s10209-017-0588-y

Coates, H. (2016). Assessing student learning outcomes internationally: insights and frontiers. *Assessment and Evaluation in Higher Education*, *41*(5), 662–676. https://doi.org/10.1080/02602938.2016.1160273

Díaz-Ramírez, J. (2020). Gamification in engineering education – An empirical assessment on learning and game performance. *Heliyon*, *6*(9), e04972. https://doi.org/10.1016/j.heliyon.2020.e04972

Elizondo-Garcia, J., Schunn, C., & Gallardo, K. (2019). Quality of peer feedback in relation to instructional design: A comparative study in energy and sustainability MOOCs. *International Journal of Instruction*, *12*(1), 1025–1040. https://doi.org/10.29333/iji.2019.12166a

Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment: 'Breakthrough? Or buncombe and ballyhoo?' *Journal of Computer Assisted Learning*, *37*(5), 1207–1216. https://doi.org/10.1111/jcal.12577

González, M., Benchoff, E., Huapaya, C., y Remón, C. (2017). Aprendizaje adaptativo: un caso de evaluación personalizada. *TE & ET: Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, *19*, 65–72.

Google Trends. (2021). https://trends.google.es/trends/?geo=MX

Gupta, A., Cecil, J., Pirela-Cruz, M., & Ramanathan, P. (2019). A Virtual Reality Enhanced Cyber-Human Framework for Orthopedic Surgical Training. *IEEE Systems Journal*, *13*(3), 3501–3512. https://doi.org/10.1109/JSYST.2019.2896061

Gupta, S., Ojeh, N., Majumder, M. A., Singh, K., & Adams, O. (2021). Use of an adaptive e-learning platform as a formative assessment tool in the cardiovascular system course component of an MBBS programme [letter]. *Advances in Medical Education and Practice*, *12*, 111–112. https://doi.org/10.2147/AMEP.S299759

Handoko, W., Mizkat, E., Nasution, A., Hambali, & Eska, J. (2021). Gamification in Learning using Quizizz Application as Assessment Tools. *Journal of Physics: Conference Series*, *1783*(1). https://doi.org/10.1088/1742-6596/1783/1/012111

Heitkamp, S. J., Rüttermann, S., & Gerhardt-Szép, S. (2018). Work shadowing in dental teaching practices: Evaluation results of a collaborative study between university and general dental practices. *BMC Medical Education*, *18*(1), 1–14. https://doi.org/10.1186/s12909-018-1220-4

Hudha, M. N., Aji, S. D., & Huda, C. (2018). E-Rubric: Scientific Work Based on Android for Experimental Physic. *IOP Conference Series: Materials Science and Engineering*, *288*(1). https://doi.org/10.1088/1757-899X/288/1/012100

Keinänen, M., Ursin, J., & Nissinen, K. (2018). How to measure students' innovation competences in higher education: Evaluation of an assessment tool in authentic learning environments. *Studies in Educational Evaluation*, *58*(October 2017), 30–36. https://doi.org/10.1016/j.stueduc.2018.05.007

Maluenda Albornoz, J., Varas Contreras, M., y Chacano Osses, D. (2021). Efectos del aula invertida y la evaluación auténtica en el aprendizaje de la matemática universitaria en estudiantes de primer año de ingeniería. *Educación*, *30*(58), 206–227. https://doi.org/10.18800/educacion.202101.010

Martín-Sómer, M., Moreira, J., & Casado, C. (2021). Use of Kahoot! to keep students' motivation during online classes in the lockdown period caused by Covid 19. *Education for Chemical Engineers*, *36*(March), 154–159. https://doi.org/10.1016/j.ece.2021.05.005

Martínez, J., López, G., Guevara, I., Luna, V., y Landa, Á. (2017). La innovación educativa y la evaluación en línea como fortalecimiento de la enseñanza superior. *Revista Electrónica Sobre Cuerpos Académicos y Grupos de Investigación en Iberoamérica. 4*(8).

Miguel, J. (2020). La educación superior en tiempos de pandemia: una visión desde dentro del proceso formativo [Higher education in times of pandemic: a view from within the training

process]. *Revista Latinoamericana de Estudios Educativos*, *50*(ESPECIAL), 13–40. https://n9.cl/ov5w4
Molina, J. P., Valencia, A., y Gómez, F. (2016). Innovación docente en educación superior: Edublogs, evaluación formativa y aprendizaje colaborativo. *Profesorado*, *20*(2), 432–450. https://doi.org/10.30827/profesorado.v20i2.10428
Morton, D. A., & Colbert-Getz, J. M. (2017). Measuring the impact of the flipped anatomy classroom: The importance of categorizing an assessment by Bloom's taxonomy. *Anatomical Sciences Education*, *10*(2), 170–175. https://doi.org/10.1002/ase.1635
Ojennus, D. D. (2016). Assessment of learning gains in a flipped biochemistry classroom. *Biochemistry and Molecular Biology Education*, *44*(1), 20–27. https://doi.org/10.1002/bmb.20926
Prendes, M. P., y Cerdán, F. (2021). Tecnologías avanzadas para afrontar el reto de la innovación educativa. *RIED. Revista Iberoamericana de Educación a Distancia*, *24*(1), 33–53.
Ramírez, M. S., & García-Peñalvo, F. J. (2018). Co-creation and open innovation: Systematic literature review. En *Comunicar* (Vol. 26, Issue 54, pp. 9–18). https://doi.org/10.3916/C54-2018-01
Riccomini, F. E., Cirani, C. B. S., Carvalho, C. C. de, & Storopoli, J. E. (2019). Educational innovation: trends for higher education in Brazil. *International Journal of Educational Management*, *35*(3), 564–578. https://doi.org/10.1108/IJEM-07-2019-0245
Rodríguez, D. V., Mezquita, J. M. M., & Vallecillo, A. I. G. (2019). Innovative methodology based on educational gamification: Multiple-choice test evaluation with Quizizz tool. *Profesorado*, *23*(3), 363–387. https://doi.org/10.30827/profesorado.v23i3.11232
Sánchez, I. R. A. (2016). The Information Society, Knowledge Society and Learning Society. Referring to their training. *Bibliotecas. Anales de Investigación (Cuba)*, *24*(2), 239–231.
Sekeroglu, B., Demililer, K., & Tuncal, K. (2019). Artificial Intelligence in Education: application in student performance evaluation. *Revista Dilemas Contemporáneos: Educación, Política y Valores.*, *8*(5), 55. https://www.dilemascontemporaneoseducacionpoliticayvalores.com/index.php/dilemas/article/view/1594
Sridharan, B., Tai, J., & Boud, D. (2019). Does the use of summative peer assessment in collaborative group work inhibit good judgement? *Higher Education*, *77*(5), 853–870. https://doi.org/10.1007/s10734-018-0305-7
Tempelaar, D. (2020). Supporting the less-adaptive student: the role of learning analytics, formative assessment and blended learning. *Assessment and Evaluation in Higher Education*, *45*(4), 579–593. https://doi.org/10.1080/02602938.2019.1677855

Capítulo 43

CRITERIOS E INDICADORES PARA ORIENTAR LA EVALUACIÓN DE LA INNOVACIÓN EDUCATIVA EN LA UNAM

Alan Kristian Hernández Romo, Roberto Santos Solórzano, Verónica Luna de la Luz

"Al contrario –prosiguió Tweedledee–, si era así, podía serlo;
y si fuera así lo sería; pero como no lo es, no lo es.
Es totalmente lógico."
Lewis Carroll

INTRODUCCIÓN

La pandemia de COVID-19, propició que docentes y estudiantes de todo el mundo modificarán sus actividades académicas; independientemente del nivel educativo al que pertenecieran, fue necesario retomar y adaptar estrategias de enseñanza y aprendizaje que provenían o que se consideraban propias de la educación a distancia[1], de igual manera se requirió utilizar diversos recursos tecnológicos para la educación en línea[2], lo anterior, con el fin de dar continuidad a sus actividades escolares a pesar de las circunstancias sanitarias. En un principio todos estos ajustes fueron definidos como "Enseñanza Remota de Emergencia" (Hodges et al., 2020) debido a que fueron estrategias que contaron con poco tiempo de planificación y en su momento se consideró que iban a ser alternativas temporales. Sin embargo, después de dos años del inicio de la pandemia, sus efectos aún tienen impacto en distintos ámbitos de la vida cotidiana de las personas. En particular, en el ámbito educativo ha sido paulatino el retorno a actividades presenciales y se han implementado diversas estrategias para adaptarse a las circunstancias emergentes, en algunos casos se ha recurrido a la modalidad mixta en la que se articulan actividades de la modalidad presencial y a distancia, por lo tanto, los distintos actores educativos, entre ellos el profesorado, se ha dado a la tarea de planear, diseñar e implementar acciones y estrategias que combinan recursos y actividades de distintas modalidades, lo que probablemente ha implicado el diseño, desarrollo e incluso la implementación de distintas innovaciones educativas que requieren ser valoradas con el propósito de identificar su impacto y con ello mejorar los procesos de enseñanza y aprendizaje del estudiantado en este entorno.

1 https://recrea.cuaieed.unam.mx/glosario

2 https://recrea.cuaieed.unam.mx/glosario

Es en este contexto que se desarrolla el presente capítulo, cuyo objetivo es ofrecer un panorama general del significado de la innovación educativa, a la vez de profundizar en algunos conceptos clave para entender y plantear una primera aproximación a la valoración de innovaciones educativas en el ámbito didáctico que orienten la ardua, compleja y enriquecedora labor de los docentes. En la primera sección se abordan aspectos conceptuales y teóricos relacionados con la innovación en educación, que son el marco de referencia para, en la segunda sección, proponer algunos criterios e indicadores que guíen la reflexión sobre propuestas, estrategias, metodologías o proyectos considerados como innovadores en el ámbito de la práctica docente en distintas disciplinas.

Esperamos que este capítulo sea del interés de los lectores y motive una constante reflexión sobre las decisiones que toman los docentes en su práctica, desde las adaptaciones, ajustes o cambios, siempre en beneficio del aprendizaje del estudiantado.

PANORAMA GENERAL DE LA INNOVACIÓN EDUCATIVA

Estimado lector, si usted cuenta con la dicha de ser docente, es muy probable que al inicio de la pandemia estuvieran en su mente, entre otras inquietudes ¿qué iba a hacer para continuar con las clases con sus estudiantes de manera remota? y ¿cómo iba a proceder?, además de preguntarse ¿con qué recursos iba a contar para poder hacerlo? y ¿qué tan difícil iba a ser su uso?; también es casi seguro que pasó por un proceso de adaptación de contenidos y estrategias didácticas, además de reconocer y familiarizarse con distintas herramientas tecnológicas que le permitieron continuar con sus clases a distancia. Posteriormente, a más de dos años de haber vivido todos estos cambios y en medio de un nuevo proceso de retorno a las actividades académicas de manera presencial, es posible que, nuevamente, surjan una serie de interrogantes, tales como:

- ¿Los cambios o adecuaciones realizados a raíz de la pandemia se tradujeron en mejoras de los procesos de enseñanza y aprendizaje?
- ¿Tuvo algún impacto la incorporación de recursos tecnológicos en los procesos de enseñanza y aprendizaje?
- ¿Cómo puedo obtener evidencias sobre los cambios en el aprendizaje de los estudiantes?

Estas y otras preguntas detonan una reflexión que debe enriquecerse con algunos referentes teóricos y conceptuales sobre la innovación en educación y su evaluación. En este orden de ideas, es pertinente señalar que la innovación es un término polisémico, complejo, con matices y connotaciones en función del contexto cultural, científico, técnico o profesional en que se utiliza (Rivas-Navarro, 2000). Diversos autores la han definido desde distintos enfoques, (Sánchez et al., 2018). Desde el punto de vista etimológico proviene del latin *innovatio* que alude a la idea de introducir algo nuevo a un contexto determinado (Rivas-Navarro, 2000). No obstante que el significado etimológico da una noción general de la innovación, no es conveniente quedarnos solo con esa definición, ya que conforme el concepto ha evolu-

cionado, también ha sido definido desde distintas perspectivas; en este sentido, varios autores han realizado contribuciones que aportan elementos importantes a considerar cuando se evalúan innovaciones en educación.

Una de estas perspectivas describe a la innovación como un proceso que permite a las personas y organizaciones cambiar sus ideas en productos, servicios o procesos, para modificar o mejorar una realidad preexistente (Baregheh et al., 2009; Sánchez-Mendiola y Escamilla de los Santos, 2018). Este proceso está conformado por los siguientes componentes:

Tabla 1. Componentes del proceso de innovación

Componente	Definición	Ejemplos en educación
Etapas	Pasos a seguir durante el proceso de innovación.	A nivel curricular se llevan a cabo una serie de pasos para crear o actualizar un plan de estudios, ponerlo en marcha, implementarlo y evaluarlo; a nivel didáctico el profesorado planea, implementa, desarrolla y evalúa su curso.
Social	Contextos o grupos de personas involucradas en el proceso de innovación.	Puede ser una institución educativa, una escuela o facultad, un grupo de profesores o bien un grupo de estudiantes de una asignatura.
Medios	Recursos o herramientas de los que se puede echar mano para llevar a cabo innovaciones.	Uso o aplicación de una estrategia didáctica que fomente el aprendizaje activo o bien la integración de un recurso tecnológico en la enseñanza y evaluación del aprendizaje de alguna disciplina.
Naturaleza	Tipo de modificación que la innovación incorpora en un contexto.	Cambio en las prácticas docentes, mejora en la enseñanza de algún contenido o novedosas maneras de evaluar.
Tipo	Manera en que se ofrece o materializa la innovación.	Puede ser un programa computacional que favorezca el desarrollo de habilidades motrices o el desarrollo de una clase que fomente el aprendizaje activo del alumnado.
Objetivo	Resultados esperados de implementar una innovación.	Favorecer el aprendizaje significativo del estudiantado. Obtener resultados objetivos y veraces derivados de las evaluaciones del y para el aprendizaje.

Tomado de: Baregheh, A., Rowley, J. and Sambrook, S. (2009).

Como puede observarse, esta serie de componentes están presentes de una u otra manera en distintos contextos educativos y son elementos que necesitan considerarse al evaluar las innovaciones en educación.

Otro marco teórico para entender y profundizar en el concepto de innovación y conocer algunos elementos que son necesarios para analizarlo es el propuesto por el sociólogo

Everett Rogers (2003), quién describe a la innovación como "una idea, práctica u objeto que es percibido como nuevo por un individuo u otra unidad de adopción"; bajo esta premisa la percepción de novedad, cambio o mejora es relativa y está en función del contexto en el que se introduce o se lleva a cabo. En el ámbito educativo esta característica tiene un papel relevante, particularmente en la educación superior, pues es en este nivel educativo que se hacen más evidentes las diferencias entre diversos campos disciplinares, cada uno con circunstancias que influyen en la percepción, creación, implementación y evaluación de sus innovaciones.

Rogers (2003) señala que cada proceso o producto innovador se comporta distinto dependiendo de la comunidad o grupo social que lo adopta, las principales diferencias se observan en el grado y tiempo que se lleva la adopción de una innovación y sugiere que esto depende de las siguientes variables:

1) La **ventaja relativa:** se refiere al grado en el que se percibe que una innovación es algo mejor a lo que existía previamente.
2) La **compatibilidad:** se refiere al grado en el que se percibe que una innovación es congruente con los valores, experiencias y necesidades de un grupo o comunidad.
3) La **complejidad:** se refiere al grado de dificultad que se percibe cuando se adopta una innovación.
4) La **capacidad de prueba:** que se refiere a la percepción sobre la posibilidad de experimentar o probar una innovación antes de adoptarla y, finalmente,
5) La **observabilidad:** que se refiere al grado en que son visibles los beneficios de una innovación.

Estas variables constituyen otros elementos que son necesarios considerar cuando se realiza la evaluación de una innovación en educación.

Aunado a lo anterior vale la pena tener en cuenta las propuestas de algunos autores sobre los aspectos en que inciden las innovaciones en educación; por ejemplo, Valenzuela-González (2017) y Ramírez-Montoya (2020) señalan que las innovaciones tienen un impacto en los niveles organizacionales, administrativos, pedagógicos, instruccionales, así como en la formación y actualización docente, mientras que, Barraza (2005) sugiere que las innovaciones en educación pueden desarrollarse e implementarse en los ámbitos institucional, curricular y didáctico (Tabla 2).

Tabla 2. Ámbitos de la innovación educativa

Institucional	Curricular	Didáctico
Prácticas políticas: negociación del conflicto y toma de decisiones.	***Prácticas de elaborar diagnósticos:*** definición de modelos y construcción de estrategias de recolección de información.	***Prácticas de planeación didáctica:*** elaboración de registros, construcción de modelos y definición de procesos.
Prácticas administrativas: planeación, dirección, organización, comunicación y evaluación.	***Prácticas de estructuración curricular:*** definición de modelos y enfoques.	***Prácticas de intervención didáctica:*** construcción de estrategias didácticas y medios para la enseñanza.
	Prácticas de evaluación curricular: definición de modelos y construcción de estrategias de recolección de la información.	***Prácticas de evaluación de los aprendizajes:*** diseño de instrumentos y construcción de estrategias.

Tomado de Barraza (2005).

Ante esta diversidad de enfoques para definir y abordar la innovación educativa y su valoración, y considerando el propósito de este capítulo, nos centraremos en describir, analizar y sugerir algunos criterios e indicadores, que permitan principalmente la reflexión sobre la evaluación de innovaciones en la dimensión didáctica, por ser el ámbito en el que los docentes tienen un mayor grado de intervención. En este sentido, vale la pena puntualizar y describir algunos de los roles de la docencia y vincularlos con las prácticas de planeación e intervención didáctica, así como, de evaluación del aprendizaje descritas por Barraza (2005).

ROLES DE LA DOCENCIA EN EDUCACIÓN SUPERIOR

Quienes tienen la experiencia de ser docentes, saben muy bien que las actividades del profesorado no se limitan a las clases frente al grupo; la docencia implica una serie de actividades entre las que podemos destacar: la planeación, búsqueda de información, evaluación de aprendizajes, desarrollo de recursos didácticos, cuyo propósito es contribuir al aprendizaje del estudiantado. Harden y Crosby (2000) hacen una descripción de distintos roles que el profesorado de disciplinas en el área de las ciencias de la salud realiza; no obstante que estos roles se ejemplifican para la enseñanza de las ciencias médicas, podríamos extrapolarlos y decir que su aplicación es universal ya que la pueden llevar a cabo los docentes de cualquier disciplina. Estos roles se agrupan en seis actividades generales:

1) **Proveer información**. En este rol aún perdura la expectativa tradicional donde se considera que los docentes son los responsables de transmitir la información, conocimiento y comprensión de un tema, sin embargo, los autores señalan que los profesores ayudan al

alumnado a interpretar el nuevo conocimiento echando mano de distintas estrategias educativas ya sea en un salón de clases, en un entorno virtual o bien en un entorno práctico;

2) **Modelar la actividad profesional.** En educación superior, los estudiantes no solo aprenden de lo que dicen sus profesores, sino también lo hacen observando las actitudes, habilidades y conocimientos que los docentes exhiben, por ello, el modelado de roles profesionales y docentes es importante para transmitir valores, actitudes, y patrones de conducta y razonamiento en los estudiantes;
3) **Facilitar el aprendizaje.** Al tener como protagonistas del aprendizaje a los estudiantes y promover en ellos el aprendizaje activo**,** los docentes tienen un rol fundamental como facilitadores de contenidos y recursos para que los estudiantes vivan experiencias, adquieran conocimientos y desarrollen habilidades, con su acompañamiento y supervisión;
4) **Evaluar el aprendizaje.** Una de las tareas más importantes que realizan los docentes es la evaluación del aprendizaje de los estudiantes, como se describe a lo largo de este libro, la evaluación del y para el aprendizaje es una parte integral de la enseñanza y en torno a ella se han desarrollado una gran cantidad de métodos e instrumentos, algunos con un alto grado de sofisticación, con el fin de evaluar de manera válida, abierta, justa y congruente con los objetivos del curso a los estudiantes;
5) **Planificar actividades de aprendizaje.** Los docentes antes de iniciar un curso deben considerar entre otras cosas: los contenidos a abordar, los resultados de aprendizaje esperados, las distintas estrategias didácticas que se pueden emplear, los procedimientos e instrumentos de evaluación adecuados, así como, el entorno educativo donde se llevará a cabo un curso; con el propósito de diseñar las experiencias de aprendizaje más enriquecedoras para los estudiantes y prever los recursos necesarios para realizarlas;
6) **Desarrollar recursos de aprendizaje.** Con el aprendizaje centrado en los estudiantes, el profesorado tiene un rol más preponderante como proveedor de información y facilitador del aprendizaje, pero también se convierte en desarrollador de materiales y recursos educativos que comparte con sus estudiantes con el propósito de brindar información a través de distintos medios y formatos, que permitan un mejor aprendizaje por parte del alumnado.

Como se puede observar, estos roles y actividades están estrechamente vinculados con las prácticas del ámbito didáctico descrito por Barraza (2005), donde muchas veces las innovaciones educativas emergen, sin embargo, por la falta de pautas para reconocerlas y evaluarlas, estas son pasajeras y fugaces, motivo por el cual en la siguiente sección de este capítulo se señalan y describen algunos criterios e indicadores que pueden ser de utilidad para proponer formas, tiempos y medios para valorar el impacto de las innovaciones en distintos contextos educativos.

CRITERIOS E INDICADORES PARA VALORAR LA INNOVACIÓN EDUCATIVA

La comprensión de la innovación educativa y la capacidad para evaluarla es fundamental para la mejora de la educación. En el ámbito pedagógico, monitorear sistemáticamente cómo las prácticas están cambiando dentro de las aulas, cómo el profesorado se desarrolla profesionalmente, cómo usa los recursos educativos y en qué medida la innovación está vinculada al logro de mejores resultados de aprendizaje, podría orientar más claramente las intervenciones y los recursos, también permitiría obtener información rápida sobre si cambiaron las prácticas educativas como se esperaba y, comprender mejor las condiciones y el impacto de la innovación educativa (Vincent-Lancrin et al., 2019).

Sin embargo, la innovación educativa en el ámbito didáctico, al ser realizada por cada docente solo tiene un alcance local (su asignatura). Esta situación ocasiona que los resultados de la innovación sean repetitivos, no transferibles entre asignaturas distintas y con una gran divergencia de resultados, también origina que el desarrollo de la innovación sea lento, tenga un impacto transformador limitado y haya una gran dificultad en valorar lo que es o no es innovación (Fidalgo-Blanco y Sein-Echaluce, 2018; Sein-Echaluce, Fidalgo-Blanco y García-Peñalvo, 2019).

Es por ello que las innovaciones docentes requieren ser evaluadas (Gómez, Vela, Iturbe, 2015), tema que, como se pudo ver en la sección anterior, se dificulta por la propia naturaleza compleja de la innovación educativa, no obstante, es un reto que se debe asumir, ya que se trata de un proceso que permite recabar información sobre la gestión, el diseño, el desarrollo, la implementación y el impacto de la innovación educativa, así mismo, brinda elementos para emprender proyectos innovadores con mayor eficiencia y de mayor calidad (León-Guerrero, López-López, 2014).

Ante esta situación surge el desafío de diseñar y desarrollar instrumentos que permitan evaluar las nuevas propuestas para la enseñanza y el aprendizaje (Gómez, Vela, Iturbe, 2015), y que permitan dar respuesta a preguntas relevantes, por ejemplo: ¿Cómo podemos distinguir entre los proyectos innovadores de aquellos que solo presentan ajustes a una realidad existente? ¿Cuáles son los criterios que permiten valorar la innovación educativa? ¿Qué factores inciden en que algunas ideas educativas novedosas se instrumenten con éxito? Y ¿En qué medida una estrategia, producto o proceso es considerada una innovación?

Pero, el proceso para realizar y validar la innovación no suele asociarse a indicadores, planificación, previsión de resultados y gestión, entre otras causas porque la innovación educativa se considera como algo etéreo, abstracto, indefinido e imprevisible. Esto ocasiona que las y los docentes no tengan guías o pautas para planificar, aplicar y validar las innovaciones que desarrollan (Fidalgo-Blanco y Sein-Echaluce, 2018; Sein-Echaluce et al., 2020).

Esta situación no es exclusiva de los docentes, pues es común que las instituciones educativas tampoco tengan criterios para identificar la innovación educativa, ni para clasificarla, organizarla y distribuirla efectivamente, lo que supone una dificultad por parte de las universidades para llevar a cabo el reconocimiento del trabajo de innovación realizado por su profesorado (Fidalgo-Blanco y Sein-Echaluce, 2018; Sein-Echaluce et al., 2020).

En el caso de la Universidad Nacional Autónoma de México (UNAM), el panorama se complica aún más, entre otras cosas, por el tamaño y la diversidad de áreas, disciplinas y niveles educativos que comprende su trabajo académico y su oferta formativa que contempla tres planes de bachillerato (nivel medio superior), 131 licenciaturas (nivel superior), 42 programas de posgrado y 42 de especialización, impartidos por una numerosa planta académica de 41,542 docentes, que atiende a una matrícula estudiantil inmensa de 366,930 estudiantes, en una gran cantidad de entidades que comprende: 14 planteles de educación media superior, 15 facultades, 9 escuelas nacionales, 35 institutos, 13 centros de investigación, 5 unidades multidisciplinarias, ubicadas en las 32 entidades federativas de México además de otros nueve países (UNAM, 2021).

Ante este escenario la Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia (CUAIEED) se ha propuesto el desarrollo de un instrumento que contenga criterios e indicadores que permitan evaluar los proyectos de innovación educativa, mismos que puedan servir como un referente y, al mismo tiempo, sean susceptibles de ser aplicados en la multiplicidad de contextos educativos que comprende esta casa de estudios, con miras de avanzar en la construcción de una visión institucional de este proceso. Hasta este momento se han planteado una serie de criterios e indicadores obtenidos a partir de la recuperación, revisión y análisis de publicaciones sobre la evaluación de la innovación educativa de distintas instituciones educativas de nivel superior nacionales e internacionales (Santos, 2021).

Originalmente se presentó como una propuesta en dos niveles; en el primero, se establecieron siete criterios, definidos como un conjunto de características o aspectos centrales que definen a la innovación educativa, dichos criterios corresponden a los componentes inherentes y que no pueden excluirse de la innovación educativa, estos son: i) novedad, ii) intencionalidad, iii) implementación, iv) impacto, v) contextualidad, vi) sostenibilidad y vii) reflexión; mientras que en el segundo nivel, para cada criterio se agregan una serie de indicadores, definidos como las unidades explícitas y verificables que permiten delimitar aspectos específicos de los criterios, es decir, sentencias que nos permiten observar elementos propios de cada uno de los criterios establecidos en el primer nivel (Santos, 2021).

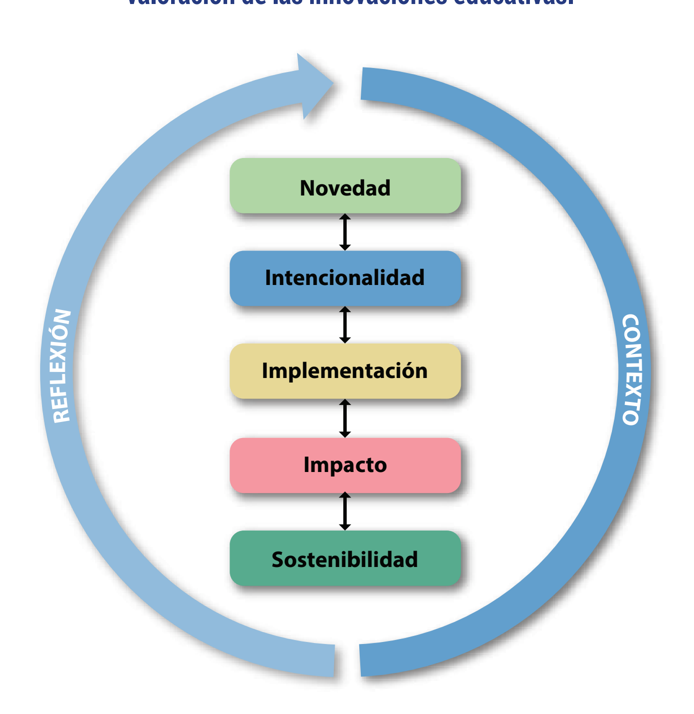
Si bien por separado y de manera individual los criterios que proponemos nos ayudan a entender y a profundizar en la complejidad de la innovación educativa, bajo la idea que, si conocemos los engranajes de una máquina, eventualmente vamos a entender cómo funciona, estos mismos criterios en conjunto adquieren otra dimensión, ya que podemos observar que la innovación educativa es mucho más que solo la suma de sus partes.

Valga la referencia y analogía con la biología desde donde se señala que un ser vivo no puede explicarse en su totalidad a partir de solo conocer sus constituyentes de manera individual, ya que muchas de sus propiedades emergen a partir de las interacciones entre los elementos que los componen, hecho conocido como *propiedades emergentes* (Acerca Ciencia, 2021).

Del mismo modo podemos observar que cuando los distintos componentes de la innovación educativa se relacionan unos con otros adquieren nuevos atributos que le dan sentido a la propuesta; por ejemplo, para diseñar y evaluar una iniciativa de innovación educativa,

no basta solo con identificar y valorar la estructura de sus elementos, sino también se requerirá describir el contexto en el que se lleva a cabo y la reflexión que la acompaña (Rivas-Navarro, 2000; Rogers, 2003), en este sentido podemos decir que los criterios de contextualidad y reflexión son los ejes que nos permiten entender cada uno de los aspectos dentro de una innovación, además, facilitan la toma de decisiones y le dan sentido a todo su desarrollo (Figura 1).

Figura 1. Los criterios de contexto y de reflexión como ejes en el proceso de valoración de las innovaciones educativas.



Elaboración propia a partir de: Santos, R. (2021). Evaluating educational innovation at UNAM: a proposal of criteria and indicators. 2021 XVI Latin American Conference on Learning Technologies (LACLO), pp. 482-485. https://doi.org/10.1109/LACLO54177.2021.00084

Los criterios que aquí se presentan pueden ser referentes para el diseño de iniciativas innovadoras como experiencias de aprendizaje, materiales y recursos educativa, o bien, pueden ser utilizados como guía para valorar internamente aquellos proyectos que ya estén en desarrollo. Esta propuesta, describe siete criterios relacionados con la innovación educativa y para cada criterio hay relacionados una serie de indicadores que orientarán al lector en el proceso de valoración de su innovación, se sugiere que en cada indicador se utilice una escala de respuesta dicotómica (presencia/ausencia; ocurrencia/no ocurrencia). Es importante enfatizar que la intención no es determinar cuáles proyectos son innovadores, sino, identificar aquellos aspectos que están más consolidados y cuáles requieren más atención para realizar los ajustes pertinentes a partir de analizar cada uno de los criterios.

El primer criterio es la ***novedad***, concepto inherente a toda propuesta de innovación. Responde a las preguntas ¿Ya se había desarrollado antes? ¿Qué tan original es? ¿Qué es lo diferente o nuevo en mi propuesta? Pero es importante que tomemos en cuenta que, como ya se mencionó antes, la innovación educativa es relativa a un contexto determinado, a un tiempo y un lugar específicos, donde generalmente ya existen antecedentes que se aprovechan como insumo para generar nuevas propuestas de solución, de este modo una innovación no necesariamente debe ser una invención o algo nunca antes visto, más bien se puede tratar de una manera diferente de configurar y vincular los elementos ya existentes, una innovación busca introducir algo nuevo en una situación dada y propiciar una mejora respecto a una situación bien definida (Ortega Cuenca et al., 2007, Rivas-Navarro, 2000). Para orientarse en la reflexión sobre este criterio sírvase de la siguiente Tabla:

Tabla 3. Indicadores para valorar el criterio de novedad

Criterios	Indicadores	Ocurrencia / ausencia
Novedad	Busca transformar un componente del proceso educativo.	•
	Introduce un enfoque, producto o estrategia, que es concebido como nuevo u original para su contexto.	•
	La propuesta es actualizada, es decir, aún es pertinente para el problema planteado.	•
	Permite que los involucrados mantengan una actitud abierta a nuevos cambios.	•
	La propuesta es creativa en su diseño, desarrollo, implementación o en algún otro aspecto, es decir, genera o adapta ideas novedosas, introduciendo elementos originales frente al desafío del ideal educativo o de una problemática educativa priorizada.	•

El segundo aspecto es la ***intencionalidad,*** este se refiere a que la innovación debe tener un carácter deliberado, es decir, que debe desarrollarse de manera consciente, voluntaria e intencional después de un proceso de análisis y reflexión. Permite responder a las preguntas: ¿Cuál es la meta de la innovación? ¿Qué problema educativo se busca atender? ¿Por qué se decidió desarrollar esta propuesta? Es importante señalar que sin una intención y una planeación explícitas que propongan lograr una mejora congruente con la institución, sus programas y planes de estudio no se pueden considerar como una innovación (Ortega Cuenca et al., 2007; Rivas-Navarro, 2000).

Tabla 4. Indicadores para valorar el criterio de intencionalidad

Criterios	Indicadores	Ocurrencia / ausencia
Intencionalidad	Identifica claramente el problema educativo al que se pretende atender.	•
	Establece objetivos y/o metas claras y congruentes al problema.	•
	Considera riesgos, conflictos, limitaciones y/o necesidades especiales.	•
	Pretende cambios relevantes o significativos en algún proceso educativo, nivel organizativo o curricular, o bien, cambios en las concepciones, las actitudes y las prácticas educativas.	•

En siguiente lugar tenemos a la ***implementación***, que se refiere a que toda innovación debe estar orientada a los resultados, por lo que se requiere informar y describir cómo se instrumentará. Este criterio permite responder a preguntas como: ¿Cuáles son las estrategias o metodologías pertinentes para la implementación? ¿Qué tipo y qué características deben tener los espacios y los medios para instrumentar la innovación? ¿Se requiere algún requerimiento especial? No hay que perder de vista que la finalidad del diseño de una innovación no es la innovación en sí misma, sino encontrar una forma de lograr de una mejor manera los fines educativos (FONDEP, 2014).

Tabla 5. Indicadores para valorar el criterio de implementación

Criterios	Indicadores	Ocurrencia / ausencia
Implementación	Se instrumentó en la enseñanza de alguna asignatura con estudiantes y presenta evidencias, contribución, análisis y sistematización.	•
	Si no se instrumentó se explica ¿por qué? y qué se requiere para hacerlo.	•
	Describe estrategias o metodologías pertinentes, congruentes entre los distintos actores.	•
	Es viable, sin la necesidad de elementos o recursos agregados que sobrepasan las capacidades del docente y/o la institución.	•
	Describe espacios y mecanismos para la participación (formales y no formales), es decir, procesos a través de los cuales los actores de la experiencia proponen, se apropian y legitiman el cambio, participando activamente en la toma de decisiones y su implementación.	•

El siguiente criterio a considerar es el ***impacto***, que se refiere a los cambios y efectos generados en los participantes o población beneficiaria de la innovación. Permite responder a preguntas como: ¿En qué medida se generó un cambio? ¿Qué tanto se ha avanzado en el logro de objetivos? ¿Cómo se medirá la experiencia? Sin olvidar que la innovación busca una auténtica transformación, una ruptura en las estructuras en el funcionamiento rutinario de los procesos de enseñanza y aprendizaje. Sin esta transformación auténtica, solo podríamos hablar de ajustes, pero no de innovación (FONDEP, 2014, López y Heredia, 2017).

Tabla 6. Indicadores para valorar el criterio de impacto

Criterios	Indicadores	Ocurrencia / ausencia
Impacto	Contempla acciones e instrumentos de evaluación de la experiencia, es decir, que permite validar y valorar el efecto de la innovación.	•
	Señala la población beneficiaria: características y cantidad.	•
	Muestra los resultados obtenidos en la población a la que va dirigida, de modo que permite responder ¿Cuál fue su aportación?	•
	Cuenta con procesos de sistematización, investigación y comunicación de la experiencia.	•
	Potencia el logro de objetivos o metas de aprendizaje.	•

Otro aspecto muy importante es el de ***contextualidad***, que parte de la idea que el contexto es uno de los factores más importantes en el éxito de las innovaciones, por lo que se deben tomar en cuenta sus características, de tal modo que la mejora que representa la innovación debe responder a los intereses de la población a la que se suscribe, una situación dada, un entorno cultural, histórico o cualquier otro, para que su impacto sea significativo y los cambios que se produzcan tengan cierta permanencia. Este criterio permite responder a preguntas como: ¿Se están atendiendo las necesidades educativas de nuestra dependencia? ¿Tiene congruencia con el modelo educativo de la Universidad? ¿La propuesta se ajusta al nivel educativo y características de la población beneficiaria? De este modo recordemos que el contexto es tan determinante que, lo que en uno puede ser innovador, no necesariamente lo será en el otro y en la medida que nuestra población reconozca como valiosa la propuesta mayor será la apropiación (FONDEP, 2014, López y Heredia, 2017, Ortega Cuenca et al., 2007, Rivas-Navarro; Rogers, 2003).

Tabla 7. Indicadores para valorar el criterio de contextualidad

Criterios	Indicadores	Ocurrencia / ausencia
Contextualidad	Responde a las necesidades o demandas de su contexto.	•
	Es compatible con los valores de quienes deben aceptarla o adoptarla, es decir, define el grado en el cual la innovación armoniza con las necesidades y con los actores involucrados.	•
	Se ajusta, se adecua o tiene una clara relación con los objetivos y el modelo educativo institucional y con el contexto sociodemográfico.	•
	Considera las condiciones de infraestructura y tecnología de su entorno.	•
	Incorpora elementos de su contexto.	•

Un siguiente criterio es el de ***sostenibilidad***, que señala que se requiere contemplar procesos y acciones que generen las condiciones para mantener la propuesta activa el tiempo necesario, con la finalidad de que los cambios propuestos se conviertan en la nueva normalidad, de darle la oportunidad a la innovación de crecer, extenderse y poder trasladarse a otros equipos y contextos. Este criterio nos permite responder a preguntas como: ¿Qué requiere mi innovación para sostenerse el tiempo necesario? ¿Qué tanta dependencia tiene mi propuesta de los recursos utilizados? ¿Qué se necesita para que otras personas, equipos o dependencias también puedan utilizarla? Y es que, aunque la innovación surge en un contexto y momento específico, su tiempo de aplicación no puede ser tan corto que señale improvisación, debe permanecer en un largo plazo (FONDEP, 2014, López y Heredia, 2017).

Tabla 8. Indicadores para valorar el criterio de sostenibilidad

Criterios	Indicadores	Ocurrencia / ausencia
Sostenibilidad	Menciona aspectos o acciones que permiten dar seguimiento y continuidad al proyecto.	•
	Aunque surge en el contexto de una disciplina o área, puede impactar más allá de ésta.	•
	Contempla recursos financieros y materiales.	•
	Promueve una gestión que permite transferir el desarrollo a otros equipos.	•
	Brinda información sobre la potencial obsolescencia, desuso, desactualización, caducidad.	•
	Tiene el potencial para convertirse en una actividad permanente.	•
	Tiene la capacidad para promover su mejoramiento.	•
	Potencialmente se permite su adopción, difusión, alcance, interdisciplina, modificaciones, mantenimiento, ampliación, interoperabilidad.	•

El último criterio es el de ***reflexión***, que se refiere a que la innovación es una acción planeada que involucra procesos de reflexión crítica no solo de la práctica sino también de la innovación misma. Este criterio permite responder preguntas como: ¿Cuál es la fundamentación de la propuesta? ¿Se contemplan aspectos éticos? ¿Cuál es la mejor manera de evaluar la propuesta? La innovación educativa es mucho más que solo una buena práctica, consiste en un proceso constante y sistemático de observar y analizar críticamente los procesos de enseñanza y aprendizaje, de modo que la propuesta, el diseño y la implementación deben estar basadas en las mejores evidencias, así como en la experiencia, y debe considerar la investigación, sistematización, formación, comunicación y evaluación de la experiencia innovadora (FONDEP, 2014, Ortega Cuenca et al., 2007, Rivas-Navarro, 2000).

Tabla 9. Indicadores para valorar el criterio de reflexión

Criterios	Indicadores	Ocurrencia / ausencia
Reflexión	Se fundamenta en conocimientos y evidencias. Es decir, surge de procesos de sistematización de información, investigación, evaluación y análisis.	•
	Utiliza acciones de evaluación para la toma de decisiones.	•
	Considera aspectos éticos.	•
	Se aprecia una observación crítica.	•

CONCLUSIONES

La innovación educativa ha tenido gran relevancia en las últimas décadas, de manera que ha sido un término frecuente en el discurso educativo, no obstante, tiene una naturaleza compleja, en primer lugar, se trata de un concepto polisémico que ha cambiado a lo largo del tiempo y que ha adquirido distintos significados de acuerdo a la disciplina, el propósito y el enfoque, además, la innovación educativa puede referirse a productos, servicios o procesos.

Otra de sus complejidades, viene de la mano con que adquiere distintas dimensiones y por ende, distintas características de acuerdo al ámbito donde se lleve a cabo, por ejemplo, cuando hablamos del ámbito docente, los actores principales son las y los docentes y sus estudiantes, en ese ámbito uno de los propósitos más importantes es mejorar los procesos de enseñanza y de aprendizaje, a partir de la implementación o adaptación de metodologías y técnicas didácticas novedosas, la incorporación de uno o varios recursos y herramientas, la búsqueda de diversas formas de evaluación del y para el aprendizaje, por mencionar algunos.

Sin embargo, para que la innovación educativa en el ámbito docente cumpla con sus objetivos requiere pasar por un proceso de evaluación, situación muy compleja, por la propia naturaleza de la innovación ya señalada, que confiere una gran variedad de retos y obstáculos para su implementación. Para orientar este proceso de evaluación, proponemos reflexionar sobre siete criterios o aspectos relacionados con la innovación educativa: i) la novedad, ii) la intencionalidad, iii) la implementación, iv) el impacto, v) la contextualidad, vi) la sostenibilidad y, vii) la reflexión.

En todas las universidades, incluyendo a la UNAM, existe la necesidad de promover y evaluar las iniciativas de innovación educativa dentro de su institución, no obstante, esta responsabilidad no recae única y exclusivamente en las autoridades, administradores y directivos, si no también, en otros muchos actores, entre los que se encuentra su profesorado. Si bien es cierto que a nivel institucional se deben generar las condiciones para promoverla, gestionarla, difundirla e incluso, propiciar que las innovaciones se integren como la nueva normalidad, en el ámbito didáctico, la acción de los docentes es crítica para

ponerlas a prueba, obtener información relevante y mejorar los procesos de enseñanza y de aprendizaje.

Esta serie de criterios e indicadores funcionan como una brújula que orienta con mayor claridad las acciones docentes cuando se ha decidido iniciar un proceso de innovar en la enseñanza, son directrices para el diseño, la gestión, la organización y, por supuesto, la evaluación del acto de innovar. Incluso, es ideal que el profesorado utilice estos criterios como una herramienta para evaluar sus proyectos en desarrollo, y como una guía para la fase de diseño. Si bien se recomienda que se atienda cada criterio por separado para conocer a detalle e identificar aquellos aspectos que están mejor consolidados y en cuales se requiere poner más atención para realizar los ajustes pertinentes, el proceso de análisis debe considerar el conjunto de criterios como un todo, ya que como se mencionó, la innovación educativa es mucho más que solo la suma de sus partes.

El diseño e implementación de un instrumento para evaluar innovaciones en educación, requiere un arduo trabajo, debe ser planificado y sistematizado, por lo tanto, la definición de esta serie de criterios e indicadores es un paso importante en el proceso encaminado a construir un instrumento de evaluación de innovaciones en la enseñanza, que permita impulsar la generación de innovaciones auténticas y duraderas y que reemplace la visión del cambio por el cambio, por una mirada en la que se contemplen las acciones planificadas que permitan introducir variaciones en el proceso de enseñar como resultado de una evaluación constante.

REFERENCIAS

Acerca Ciencia (15 diciembre, 2021). Más que la suma de las partes. https://www.acercaciencia.com/2019/09/30/mas-que-la-suma-de-las-partes/

Baregheh, A., Rowley, J. and Sambrook, S. (2009), “Towards a multidisciplinary definition of innovation”, Management Decision, Vol. 47 No. 8, pp. 1323-1339. https://doi.org/10.1108/00251740910984578

Barraza Macías, A. (2005). Una conceptualización comprehensiva de la innovación educativa. *Innovación Educativa*, *5*(28).

Harden R.M., Crosby J. (2000). AMEE Guide No 20: The good teacher is more than a lecturer - the twelve roles of the teacher. *Medical Teacher,* 22:4, 334-347, DOI: https://doi.org/10.1080/014215900409429

Gómez R, Vela I, Iturbe E. (2015). Categorías de evaluación de innovación educativa para mejorar la competitividad en educación superior. Mem del IX Congr la Red Int Investig en Compet. 1448–64. https://bit.ly/3mziSW2

Fidalgo-Blanco, Ángel, y Sein-Echaluce, M. L. (2018). Método MAIN para planificar, aplicar y divulgar la innovación educativa. *Education in the Knowledge Society (EKS)*, *19*(2), 83–101. https://doi.org/10.14201/eks201819283101

FONDEP (2014). *En el corazón de la escuela palpita la innovación.* Una propuesta para aprender a sistematizar experiencias de innovación y buenas prácticas educativas. Fondo Nacional de Desarrollo de la Educación Peruana. Lima, Perú.

Hodges, C., Moore, S., Lockee, B., Trust, T., Bond, A. (2020). The Difference Between Emergency Remote Teaching and Online Learning. EDUCASE Review. https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

León-Guerrero MJ, López-López MDC. (2014). Criterios para la Evaluación de los Proyectos de Innovación Docente Universitarios. *Estud. Sobre Educ,* 26, 79–101.

López Cruz, C. S. y Heredia Escorza, Y. (2017). *Escala i. Marco de referencia para la evaluación de proyectos de innovación educativa*. Guía de aplicación. Instituto Tecnológico y de Estudios Superiores de Monterrey, México.

Ortega Cuenca, P., Ramírez Solís, M. E., Torres Guerrero, J. L., López Rayón, A. E., Yacapantli Servín Martínez, C., Suárez Téllez, L., y Ruiz Hernández, B. (2007). Modelo de innovación educativa. un marco para la formación y el desarrollo de una cultura de la innovación. *RIED. Revista Iberoamericana De Educación a Distancia*, *10*(1), 145–173. https://doi.org/10.5944/ried.1.10.1023

Ramírez-Montoy M.S. (2020). Challenges for Open Education with Educational Innovation: A Systematic Literature Review. *Sustainability*, *12*(17):7053. https://doi.org/10.3390/su12177053

Rivas-Navarro M. (2000). Innovación educativa: teoría, procesos y estrategias. Edit. Síntesis, Madrid, España.

Rogers, E. M. (2010). Diffusion of innovations. Simon and Schuster.

Sánchez-Mendiola, M., y Escamilla de los Santos, J. (2018). Perspectivas de la Innovación Educativa en Universidades de México. Retrieved from https://cuaieed.unam.mx/descargas/investigacion/Perspectivas-de-la-innovacion-educativa-en-universidades-de-Mexico.pdf

Santos, R. (2021). Evaluating educational innovation at UNAM: a proposal of criteria and indicators. 2021 XVI Latin American Conference on Learning Technologies (LACLO), pp. 482-485, https://doi.org/10.1109/LACLO54177.2021.00084

Sein-Echaluce, M. L., Fidalgo-Blanco, Á., y García-Peñalvo, F. J. (2019). Diseño de un proyecto de innovación educativa docente a partir de indicadores transferibles entre distintos contextos. En M. L. Sein-Echaluce Lacleta, Á. Fidalgo-Blanco, y F. J. García-Peñalvo (Eds.), *Actas del V Congreso Internacional sobre Aprendizaje, Innovación y Competitividad. CINAIC 2019* (9-11 de octubre de 2019, Madrid, España) (pp. 617-622). Zaragoza, Spain: Servicio de Publicaciones Universidad de Zaragoza. doi: https://doi.org/10.26754/CINAIC.2019.0126

Sein-Echaluce M.L., Fidalgo-Blanco Á., García-Peñalvo F.J., Balbín A.M. (2020). Global Impact of Local Educational Innovation. In: Zaphiris P., Ioannou A. (eds) *Learning and Collaboration Technologies. Designing, Developing and Deploying Learning Experiences. HCII 2020*. Lecture Notes in Computer Science, vol 12205. Springer, Cham. https://doi.org/10.1007/978-3-030-50513-4_39

Valenzuela-González J.R. (2017). La innovación como objeto de investigación: problemas, tensiones y experiencias. En: Ramírez-Montoya M.S y Valenzuela-González J.R. (Eds.) *Innovación educativa: investigación, formación, vinculación y visibilidad*. Edit. Síntesis, Madrid, España.

Vincent-Lancrin, S., Urgel, J. Kar, S. y Jacotin, G. (2019). *Measuring Innovation in Education 2019: What Has Changed in the Classroom?*, Educational Research and Innovation, OECD Publishing, Paris, https://doi.org/10.1787/9789264311671-en

Capítulo 44

ANALÍTICAS DEL APRENDIZAJE Y EVALUACIÓN: OPORTUNIDADES PARA LA MEJORA CONTINUA CURRICULAR EN EDUCACIÓN SUPERIOR

Isabel Hilliger

INTRODUCCIÓN

La evaluación es el acto de determinar el mérito o valor de un aspecto observado, siendo aplicable a distintos ámbitos del quehacer humano. En un ámbito educativo, la evaluación es un proceso permanente, cuya finalidad es recolectar y analizar información sobre los logros de aprendizaje de las y los estudiantes (Fernández-Marcha, 2009). Indistinto del énfasis puesto en la medición de los aprendizajes, la evaluación educativa puede tener múltiples objetos de valoración. Estos pueden ser concretos, tales como la medición directa del logro de competencias o de la adquisición de conocimientos, o complejos, tales como la medición indirecta de la efectividad de recursos pedagógicos o de la percepción del estudiantado sobre la docencia. Por consiguiente, la evaluación no es un hecho puntual, sino un conjunto de tareas que se condicionan mutuamente con las actividades de enseñanza y aprendizaje.

Más aún, la evaluación educativa debe dar pie para la realización de reajustes en las actividades de enseñanza y aprendizaje, existiendo una fuerte relación entre la evaluación y la mejora continua de la calidad educativa. Por mejora continua, hacemos alusión al proceso sistemático de medir los resultados de aprendizaje en cursos específicos de un plan de estudio, con el fin de evaluar la efectividad de prácticas micro y macro curriculares (Hilliger, Celis, y Pérez-Sanagustín, 2020). Esto implica ciclos de trabajo iterativos –basados en distintos tipos de evaluaciones– en los cuales se planifican acciones para mejorar los logros de aprendizaje. A nivel de asignatura, estas acciones pueden implicar cambios en las metodologías de enseñanza o en las estrategias de evaluación, mientras que, a nivel de plan de estudio, estas acciones pueden conllevar a la revisión de la declaración y tributación de competencias en el perfil de egreso. La ejecución de estas acciones debe ocurrir en un período académico

específico, y posterior a su implementación, se debe recolectar información para analizar si se lograron los objetivos esperados.

En la educación superior, la mejora continua curricular se ha vuelto crucial para crear una cultura compartida de calidad, promoviendo la incorporación de nuevas prácticas docentes que sean relevantes y significativas para los aprendizajes (Dicker et al. 2019). No obstante, los procesos de mejora continua curricular requieren el análisis de grandes cantidades de datos, incluyendo la lista de competencias descritas en el perfil de egreso, su alineación con los resultados de aprendizaje a nivel de curso, y el nivel de logro alcanzado según instrumentos utilizados para su evaluación (Brown, DeMonbrun, y Teasley, 2018). Algunas instituciones de educación superior ya han comenzado a confiar en herramientas que recopilan y analizan evidencia en tiempo real de los resultados de aprendizaje, con el objetivo de visualizar niveles de logro alcanzados por sus estudiantes, e informar los cambios curriculares correspondientes (Hilliger, Miranda, Celis, y Pérez-Sanagustín, 2019). Este tipo de herramientas nacen dentro del campo interdisciplinario denominado analíticas del aprendizaje.

Algunos investigadores indican que la adopción de analíticas del aprendizaje podría mejorar la calidad de la educación superior en Latinoamérica y en otras regiones (Hilliger, Ortiz-Rojas, Pesántez-Cabrera, Scheihing, Tsai, Muñoz-Merino, Broos, et al., 2020). Sin embargo, existe evidencia limitada sobre el potencial impacto de esta línea de investigación en la medición de resultados del aprendizaje y la consecuente toma de decisiones institucionales (Viberg, Hatakka, Bälter y Mavroudi, 2018). Dada la necesidad de realizar más estudios para comprender cómo el uso de herramientas analíticas puede apoyar la mejora continua curricular, este capítulo describe el estado del arte de la analítica del aprendizaje y la analítica curricular, para luego profundizar en las oportunidades y desafíos asociados al uso de este tipo de herramientas para apoyar la evaluación de los aprendizajes (y la consecuente transformación de prácticas curriculares y docentes). Para guiar futuras gestiones e investigaciones, este capítulo además presenta consideraciones y recomendaciones para la adopción de analíticas del aprendizaje en Latinoamérica, proponiendo acciones para un contexto que va más allá de la pandemia mundial actual.

ORIGEN DE LAS ANALÍTICAS DEL APRENDIZAJE Y SURGIMIENTO DE LA ANALÍTICA CURRICULAR

La analítica del aprendizaje o *Learning Analytics* (LA, por sus siglas en inglés) es una línea de investigación en la cual confluye la educación, las matemáticas, y las ciencias de la computación (Romero y Ventura, 2020). Según la definición acuñada en el primer Congreso Internacional sobre Analítica del Aprendizaje y Conocimiento (LAK 2011 en Banff, Canadá), LA consiste en: 'la medición, recopilación, análisis y visualización de datos sobre el estudiantado y sus contextos, con el fin de comprender y optimizar el aprendizaje y los entornos en los que ocurre' (Siemens y Gasevic, 2012). Esto significa recopilar datos educativos de diferentes fuentes, tales como datos demográficos (por ejemplo, género, edad), datos académicos (por ejemplo, inscripción en cursos, asistencia a clases), e información relacionada con

el comportamiento de las y los estudiantes (por ejemplo, respuestas a cuestionarios, clics de los estudiantes en una plataforma de aprendizaje en línea o LMS). Para analizar e interpretar este tipo de datos se ha utilizado una amplia gama de técnicas, tales como algoritmos predictivos, análisis de conglomerados o *clustering*, o detección de valores atípicos (Romero y Ventura, 2020). Producto de estas técnicas, distintos actores educativos pueden medir logros de aprendizaje en tiempo real, entregar retroalimentación personalizada y automatizada a los estudiantes, e identificar quienes estén en riesgo de fallar –entre otras intervenciones (Cobo y Aguerrebere, 2018).

Originalmente, el campo de LA surgió de la creciente disponibilidad de datos educativos en las instituciones de educación superior producto de la interacción de estudiantes con ambientes virtuales o sistemas de gestión del aprendizaje (mejor conocidos como *Learning Management Systems* o LMS, por sus siglas en inglés) (Ifenthaler y Yau, 2020). Estas instituciones no estaban necesariamente aprovechando la existencia de estos datos para apoyar procesos de enseñanza y aprendizaje, por lo que el surgimiento del LA creó la oportunidad de utilizarlos para planificar intervenciones en distintos niveles, ya sea en la sala de clases, a nivel departamental, o institucional. Para documentar los resultados de estas intervenciones basadas en LA, se lanzó el *Journal of Learning Analytics* en 2014. Desde ese momento hasta ahora, el LA se ha convertido en un campo de trabajo reconocido, habiendo experimentado un aumento significativo en el número de artículos publicados según los resultados encontrados en buscadores web, tales como *Google Scholar* (Romero y Ventura, 2020). Esto ha atraído el interés de varias revistas científicas y de distintas líneas de financiamiento, además de motivar la existencia de una industria en crecimiento cuya finalidad es comercializar herramientas de LA como un servicio (Cechinel et al. 2020).

El creciente interés en LA se debe principalmente a su gran potencial para producir información novedosa en tiempo real para estudiantes, profesores y gestores. Por ejemplo, en lo que respecta a estudiantes, visualizaciones basadas en LA pueden proporcionarles información sobre su desempeño académico, mientras que otras herramientas pueden proveerles de tutorías personalizadas y comentarios automatizados (Archer y Prinsloo, 2020). Con respecto al profesorado, existen distintos tipos de estrategias basadas en LA para proporcionarles datos que les permitan reflexionar sobre la efectividad de sus prácticas docentes (Archer y Prinsloo, 2020; Wong y Li, 2020), tales como el tiempo de dedicación de sus estudiantes a distintas actividades de enseñanza y aprendizaje, los resultados de aprendizaje de sus estudiantes en distintos dominios, o incluso las necesidades de apoyo académico de distintos subgrupos según las probabilidades que tengan de reprobar su asignatura. Con relación a los gestores, tales como directores de carrera o coordinadores curriculares, distintos tipos de herramientas de LA pueden proporcionarles información para comprender por qué algunos subgrupos de estudiantes no se están desempeñando como se esperaba, y así poder asignar recursos para intervenciones de apoyo (Wong y Li, 2020). A su vez, existen herramientas de LA que pueden incluso retroalimentar la toma de decisiones sobre el plan de estudios, permitiendo visualizar la distribución del logro de competencias a lo largo de diferentes asignaturas, y el avance curricular del estudiantado a nivel de cursos inscritos y

desempeño académico. A este tipo de herramientas se les conoce como analítica curricular o *curriculum analytics*.

La analítica curricular surge como un subcampo del LA, acotando la recopilación y el análisis de datos educativos para respaldar específicamente la toma de decisiones del currículo y la mejora de la calidad educativa. Por analítica curricular o *curriculum analytics* (CA, por sus siglas en inglés), nos referimos a la recopilación, análisis y visualización de datos a nivel de programa y de asignatura con el objetivo de informar acciones de mejora macro y micro curriculares (Hilliger, Aguirre, et al., 2020). Esto implica el uso de diferentes tipos de datos educativos para impulsar mejoras, tanto en el diseño instruccional de una asignatura como en el desarrollo de un programa académico, aumentando el logro de resultados de los estudiantes a largo plazo (Greer, Molinaro, et al., 2016). Por tipos de datos, Ifenthaler y Yau (2020) distinguen entre tres perfiles: 1) el perfil del estudiante, 2) el perfil de aprendizaje, y 3) el perfil del plan de estudios. El perfil del plan de estudios incluye indicadores que reflejan el desempeño esperado y requerido a distintos niveles. A nivel de asignatura, existen indicadores respecto de las características de un curso, tales como el instructor, el identificador o número del curso, y las calificaciones obtenidas por el estudiantado en distintos tipos de estrategias de evaluación (por ejemplo, situación tipo prueba o entregable de un proyecto). A nivel de plan de estudio, existen otros indicadores relevantes, tales como el perfil de egreso, la secuencia sugerida de asignaturas, y el avance curricular de cada estudiante.

Hasta ahora, se han desarrollado algunas herramientas de CA que pueden ser útiles para promover la mejora continua curricular. La mayoría de estas herramientas están dirigidas a gestores y equipo docente (Hilliger, Aguirre, et al. 2020), proporcionándoles información sobre indicadores de riesgo (por ejemplo, tasas de retención, tiempo de finalización), además de representaciones gráficos sobre los flujos de estudiantes entre distintos cursos de un plan de estudio y entre distintos programas académicos (por ejemplo, transición de pregrado a posgrado). Algunas herramientas se han diseñado específicamente para las y los estudiantes, proporcionándoles gráficos de radar sobre su logro de competencias en términos de calificaciones, créditos cursados, y desempeño según logro de competencias (Chou et al., 2015). Más aún, estudios recientes de CA han centrado sus esfuerzos en mapear el desarrollo de habilidades a lo largo de un plan de estudios, respondiendo a una creciente preocupación con respecto al logro de competencias, la empleabilidad, y el aprendizaje permanente de quienes se gradúan de estudios superiores. Por ejemplo, algunos investigadores han propuesto métodos para explorar el impacto de inscribir ciertos cursos en un mismo semestre (Brown et al., 2018), analizando las descripciones de las asignaturas en términos de taxonomías de habilidades y niveles de progresión (Kitto et al. 2020).

Teniendo en cuenta la importancia de asegurar el logro de competencias al momento de obtener un título de educación superior, más iniciativas deberían promover la incorporación de herramientas basadas en CA para medir el desarrollo de habilidades a lo largo de distintas materias. Sin embargo, la evidencia actual sobre el impacto de herramientas basadas en CA no necesariamente da cuenta de cómo su adopción ha permitido transformar la forma en que las instituciones de educación superior abordan los procesos de enseñanza y aprendizaje

de sus estudiantes (Brown et al. 2018). A la fecha, la mayoría de los estudios se han concentrado en el diseño e implementación de distintos tipos de herramientas analíticas, mientras que solo algunos de estos esfuerzos han explorado las necesidades de distintos grupos de interés (Hilliger, Ortiz-Rojas, et al. 2020). En esta línea, la sección a continuación describe cómo la evaluación educativa podría beneficiarse de la adopción de herramientas analíticas, dando cuenta de sus potenciales beneficios para la mejora continua curricular y docente.

ANALÍTICAS DEL APRENDIZAJE PARA LA EVALUACIÓN EDUCATIVA

Según la base de datos de Scopus, 109 trabajos fueron publicados entre el 2017 y el 2021 sobre analítica del aprendizaje y evaluación.[1] Habiendo hecho una búsqueda con los términos "*learning analytics*" y "*assessment*" en títulos de publicaciones, resúmenes ejecutivos, palabras clave, encontramos 13 actas de diferentes conferencias y 96 estudios (46 artículos científicos, 40 ponencias de conferencias, 6 capítulos de libro, y 4 revisiones de literatura). De los 96 estudios, más de un 80% corresponde a trabajo sobre iniciativas en educación superior, los cuales han sido publicados principalmente en la revista *Assessment and Evaluation in Higher Education* o en una serie de libros publicados por Springer, los cuales son conocidos como *Lecture Notes in Computer Science*.

Indistinto del nivel educativo, los 96 estudios encontrados dan cuenta del uso de diferentes tipos de metodologías de investigación. Pese a que 38 de estos estudios no daban cuenta de la metodología utilizada en sus resúmenes o *abstracts*, los 58 estudios restantes si la detallan, describiendo el uso de variadas técnicas estadísticas, la realización de revisiones de literatura, el diseño e implementación de algoritmos basados en aprendizaje de máquinas, y el desarrollo de estudios carácter cualitativo, cuantitativo, y mixto. En relación a las técnicas estadísticas, los investigadores hacen alusión al uso de pareamientos por puntajes de propensión o vecinos cercanos para medir el efecto en el rendimiento de un subgrupo de estudiantes que ha recibido retroalimentación personalizada, así como también de otro tipo de análisis de conglomerados para agrupar a estudiantes según su comportamiento, ya sea respecto de su interacciones con actividades disponibles en el LMS o en ambientes virtuales (incluyendo evaluaciones), o específicamente en base a su rendimiento académico en una evaluación de carácter inicial o diagnóstica. Con respecto al uso de algoritmos basados en aprendizajes de máquinas, los estudios aluden al uso de redes neuronales artificiales para examinar el grado de interés del estudiantado, ya sea en lecturas previas a las clases o en actividades gamificadas, y evaluación cómo ese grado de interés incide en su participación en clases o en su desempeño académico. Adicionalmente, se utilizan algoritmos de *clustering* no supervisados para predecir el desempeño académico en base a la gestión del tiempo del estudiantado y las trazas identificadas en su uso del LMS.

---

[1] La búsqueda fue realizada el día 4 de noviembre del 2021, habiendo considerado como rango de búsqueda entre el 2017 y el presente.

A su vez, los 96 estudios encontrados también dan cuenta de distintos objetivos de investigación. La Figura 1 muestra cómo se distribuyen estos estudios según objetivos, observándose una predominancia de estudios sobre el comportamiento del estudiantado en ambientes virtuales de aprendizaje (13 de 96), sobre el uso de LA para la mejora continua curricular (10 de 96), y sobre la adopción de herramientas de LA a nivel institucional (10 de 96). En un menor grado, se identifican algunos estudios directamente relacionados con la evaluación educativa, ya sea con la entrega de retroalimentación automatizada y personalizada (7 de 96), la evaluación formativa (4 de 96), u otras prácticas evaluativas diversas (4 de 96). Entre estos últimos estudios, se describe el uso de algoritmos para predecir el rendimiento de estudiantes en exámenes finales a partir de calificaciones parciales, y el uso de sistemas de recomendadores para entregar retroalimentación personalizada y automatizada durante el transcurso de una asignatura. También se describe el uso de herramientas para facilitar la evaluación de pares y la autoevaluación, combinada con las evaluaciones docentes, así como también el uso de portafolios digitales para facilitar la medición de los logros de aprendizaje (y las consecuentes necesidades de apoyo).

Figura 1. Objetivos de investigación de estudios publicados en Scopus entre el 2017 y el 2020 sobre analíticas de aprendizaje y evaluación



Sobre el comportamiento del estudiantado en ambientes virtuales del aprendizaje, esto se ha visto acelerado durante la pandemia causada por el COVID-19, considerando que la educación superior es uno de los sectores que ha continuado sus actividades mediante el uso de modalidades de aprendizaje en línea. Según los hallazgos de una encuesta realizada por la Asociación Internacional de Universidades durante el año 2020 (IAU, 2020), un 67% de las instituciones de educación superior reemplazaron la enseñanza en el aula por la educación a distancia, y la mayoría implementó lo que los investigadores denominan "educación en línea de emergencia". Estudios previos a la pandemia ya daban cuenta de investigaciones para identificar factores de éxito académico en cursos masivos y abiertos en líneas (o mejor conocidos como MOOCs), así como en el interés de comprender cómo el andamiaje del aprendizaje incide en las preferencias del estudiantado y en sus probabilidades de completar cursos en esta modalidad. A partir del 2020, algunos estudios ya logran analizar el comportamiento del estudiantado en entornos virtuales de aprendizaje de una forma más integral, llegando a dar cuenta de beneficios de aprendizaje a partir de la incorporación de herramientas basadas en LA.

A pesar de las ventajas prometedoras de las estrategias de aprendizaje en línea y sus enfoques para la recopilación de datos, tales como proporcionar información sobre trazas del comportamiento del estudiantado en distintas actividades de enseñanza y aprendizaje, la rápida transición al aprendizaje digital también ha revelado y exacerbado problemas de equidad y calidad. En este contexto complejo, las instituciones de educación superior están tratando de asegurar la continuidad de estudios, sin necesariamente discutir y evaluar los cambios implementados como una respuesta rápida a las necesidades imprevistas de digitalización del currículo. He ahí la importancia de incorporar herramientas de analítica curricular; no solo para monitorear el aprendizaje a nivel de asignatura y plan de estudio, sino que también para promover estrategias de aseguramiento de la calidad en un corto y mediano plazo. En esa línea, los resultados de evaluaciones educativas pueden ser extremadamente útiles para documentar el efecto del diseño instruccional y para informar el proceso de enseñanza y aprendizaje, facilitando la entrega de retroalimentación a las y los según determinados puntajes alcanzados, así como también la identificación de estudiantes en riesgo de reprobar o desertar. También se pueden identificar elementos curriculares que requieran una revisión, incluyendo la definición de estrategias de evaluación que faciliten el análisis del aprendizaje por competencias. Un desafío particular de este tipo de modelos es obtener una calificación cuantitativa en función de la evaluación detallada de los resultados de aprendizaje, que luego permita utilizar este dato como el nivel de logro de un indicador de desempeño asociado a una competencia. Investigadores ya han dado un paso inicial en esta dirección, estableciendo un marco general para mejorar la calidad de los programas educativos en base al análisis de resultados evaluativos por parte del director de carrera y su respectivo equipo docente (Hilliger, Aguirre, et al., 2020; Hilliger, Celis, y Pérez-Sanagustín, 2020).

No obstante, la incorporación de herramientas basadas en LA y CA para apoyar procesos de evaluación educativa no es del todo trivial. De los 96 estudios encontrados en la base de datos en Scopus, se pueden encontrar investigaciones que se enfocan únicamente

en comprender los potenciales beneficios y barreras para adoptar herramientas analíticas a nivel institucional. Algunas de estas investigaciones revelan altas expectativas por parte del estudiantado y el cuerpo docente. Por una parte, las y los estudiantes esperan que diferentes herramientas analíticas les puedan proporcionar la posibilidad de autoevaluarse, recibiendo recomendaciones a partir del análisis personalizado de sus resultados de aprendizaje. Por otra parte, las y los docentes esperan utilizar la analítica del aprendizaje y curricular para orientar el diseño instruccional, para facilitar la autonomía de sus estudiantes y el fortalecimiento de sus hábitos de estudio. Sin embargo, estos estudios también dan cuenta de desafíos tecnológicos y éticos asociados a la recolección y análisis de datos educativos. Desde la perspectiva de gestores institucionales, existe una sensación de incertidumbre sobre los rendimientos de la inversión en este tipo de herramientas, así como también la carga que esto pueda significar para el cuerpo docente en relación con su distribución del tiempo. Por consiguiente, se requiere implementar aplicaciones analíticas a mayor escala, explorando como su uso para apoyar la evaluación educativa favorece la retención y empleabilidad de las y los estudiantes.

CONSIDERACIONES PARA LA MEJORA CONTINUA CURRICULAR EN LATINOAMÉRICA

En esta última década, la investigación en LA y CA ha evolucionado rápidamente en Norteamérica y Oceanía (Ifenthaler y Yau, 2020). Desde que esta línea de investigación fue creada, países anglosajones tales como Estados Unidos, Reino Unido y Australia han contribuido activamente a la producción científica del campo, incluso más que Europa continental (Cechinel et al., 2020). Se han realizado revisiones bibliográficas inclusivas para analizar la productividad de la investigación en diferentes territorios, incluyendo artículos escritos en español, portugués, y otros idiomas (Cechinel et al. 2020). Aun así, estas revisiones muestran que existe un número desproporcionado de estudios que documentan el diseño e implementación de herramientas analíticas en regiones desarrolladas, sin tener en cuenta las necesidades en otras regiones de desarrollo emergente (Hilliger, Ortiz-Rojas, et al. 2020). Teniendo en cuenta que los niveles de madurez de la adopción de LA y CA en Latinoamérica están muy por debajo de los de Europa, el programa Erasmus+ de la Comisión Europea financió un proyecto para instalar capacidades para el diseño e implementación de herramientas analíticas en la región. A esta iniciativa se le conoce como el proyecto LALA–*Learning Analytics for Latin America* (Maldonado-Mahauad et al. 2018), y en el contexto de este proyecto, se desarrollaron una serie de estudios y de herramientas analíticas para fomentar la mejora de la calidad educativa en distintos aspectos a la educación superior (Hilliger, Pérez-Sanagustín, et al., 2020), tales como la autorregulación del aprendizaje, la consejería académica, y la evaluación del logro de competencias.

Tres de los estudios realizados durante el proyecto LALA se enfocaron específicamente en la mejora continua curricular utilizando LA y CA. El primer estudio tuvo como objetivo explorar el beneficio potencial de usar LA para mejorar la calidad de los programas de educación superior ofrecidos en la región (Hilliger, Ortiz- Rojas, et al. 2020). Para cumplir con este objetivo de investigación, se utilizaron métodos mixtos bajo un diseño paralelo

convergente. Este tipo de diseño de investigación requiere recolectar y analizar información cualitativa y cuantitativa simultáneamente, con el fin de comparar los hallazgos obtenidos de cada tipo de datos por separado (Creswell, 2012). Para la recopilación de datos cualitativos y cuantitativos, adaptamos la versión en español de los protocolos de encuestas y evaluaciones cualitativas propuestas por el proyecto SHEILA (https://sheilaproject.eu/sheila-framework/). Esta adaptación consistió en una edición de lenguaje menor para adecuar algunas expresiones a Chile y Ecuador, considerando la ubicación de las cuatro universidades latinoamericanas afiliadas a LALA. Los datos cualitativos se recopilaron a partir de 3 entrevistas semiestructuradas y grupos focales, que abarcaron 37 gestores, 51 profesores, y 45 estudiantes de educación superior. Los datos cuantitativos se recopilaron a partir de encuestas respondidas por 1,884 estudiantes y 268 profesores. Los hallazgos que surgieron del análisis de datos cualitativos y cuantitativos se triangularon comparando los resultados de fuentes de datos y de las diferentes perspectivas de los actores educativos involucrados (Creswell, 2012).

Luego, el segundo estudio tuvo por objetivo comprender cómo una herramienta de CA puede apoyar procesos de mejora continua curricular (Hilliger, Aguirre, et al. 2022). Para cumplir con este objetivo, se realizó una investigación basada en diseño de dos ciclos para desarrollar una herramienta de CA. El primer ciclo consistió en una fase de promulgación para desarrollar una primera versión de la herramienta (https://youtu.be/WEZdBdyOLfk), seguida de una evaluación de su uso para apoyar tareas de evaluación de competencias y discusiones curriculares llevadas a cabo entre 2015 y 2017 en una universidad latinoamericana. Posteriormente, se aplicó un cuestionario en papel a 25 integrantes de los equipos docentes para evaluar la utilidad percibida y la usabilidad de esta primera versión de la herramienta, incluyendo preguntas de tipo abiertas y cerradas. Durante el segundo ciclo, se rediseñó esta herramienta (https://youtu.be/FPzf5NiJR8A), y luego se evaluó a través de estudio de campo con 16 gestores y 9 profesores de dos universidades latinoamericanas (incluyendo a la institución que participó durante el primer ciclo). Durante este estudio, se aplicó un cuestionario en línea para recolectar información de la perspectiva de ambos tipos de usuarios. Considerando los hallazgos de ambos ciclos, se realizó una evaluación amplia para capturar lecciones aprendidas para el diseño e implementación de las herramientas de CA para la mejora continua curricular.

Finalmente, el tercer estudio tuvo como objetivo comprender cómo involucrar a distintos grupos de interés en la mejora continua curricular, buscando identificar mecanismos que comprometen al equipo docente en una institución de educación superior con la evaluación de logro de competencias y discusiones curriculares a nivel de programa (Hilliger, I., Celis, S. y Pérez-Sanagustín, M. 2022). Para cumplir con este objetivo, se desarrolló un diseño de estudio de caso único, considerando que este tipo de diseño de investigación se utiliza para investigar un fenómeno en su contexto real y en el tiempo (Yin, 2014). Este estudio de caso se llevó a cabo en la misma universidad latinoamericana que participó en el primer y segundo ciclo del segundo estudio, y su redacción se basó en la recopilación de datos cuantitativos y cualitativos de tres tipos diferentes de fuentes de datos. En primer lugar, tres investigadores

analizaron 97 planes de evaluación de cursos (http://bit.ly/2ACZUXF), que clasificaron según cuatro categorías: (1) su nivel de detalle, (2) su grado de innovación (es decir, la inclusión de métodos de evaluación no tradicionales), (3) la inclusión de métodos de evaluación indirectos (por ejemplo, grupos focales y entrevistas), y (4) la variedad de métodos de evaluación propuestos por dicho plan. En segundo lugar, tres investigadores también analizaron 27 actas de reuniones (http://bit.ly/2DOS8Nu), las cuales tenían como objetivo identificar tres tipos de reflexión dentro de las notas de la reunión: 1) reflexiones sobre la necesidad de mejorar las prácticas docentes, 2) reflexiones sobre la necesidad de mejorar los métodos de evaluación, y 3) reflexiones sobre la necesidad de estrategias de renovación curricular. En tercer lugar, se realizaron 11 entrevistas semiestructuradas con profesores de enseñanza, y sus transcripciones textuales se analizaron a través de la codificación inductiva con NVivo Pro 12. Por consiguiente, para construir el estudio de caso, se analizaron similitudes y contrastes a partir de la evidencia obtenida de estas diferentes fuentes, identificando mecanismos que influyen en el compromiso del profesorado con las tareas de mejora continua.

Tabla 1. Principales resultados de estudios realizados en el contexto del proyecto LALA sobre el uso de analíticas del aprendizaje para la mejora continua curricular

Referencia al estudio en APA	Resultados
Estudio 1. Hilliger, I., Ortiz-Rojas, M., Pesántez-Cabrera, P., Scheihing, E., Tsai, Y. S., Muñoz-Merino, P. J., ... & Pérez-Sanagustín, M. (2020). Identifying needs for learning analytics adoption in Latin American universities: A mixed- methods approach. *The Internet and Higher Education, 45, 100726.*	• Los estudiantes necesitan retroalimentación y apoyo oportuno para mejorar sus resultados de aprendizaje, particularmente cuando enfrentan dificultades que afectan su desempeño académico. • El equipo docente necesita alertas tempranas para brindar un mejor apoyo al estudiantado que enfrenta dificultades que afectan su desempeño académico, así como retroalimentación significativa sobre su desempeño docente. • Los gestores necesitan información de calidad para evaluar las intervenciones de apoyo dirigidas al estudiantado.
Estudio 2. Hilliger, I., Aguirre, C., Miranda, C., Celis, S., & Pérez-Sanagustín, M. (2022). Lessons learned from designing a curriculum analytics tool for improving student learning and program quality. *Journal of computing in higher education*, 1-25. https://doi.org/10.1007/s12528-022-09315-4	• Los gestores y equipo docente emergen personal como los principales usuarios de una herramienta de CA. • Las herramientas de CA permiten recopilar una mayor cantidad y variedad de evidencia documental sobre el logro de competencias. • Los gestores y el equipo docente consideran que las herramientas de CA son útiles para asegurar los aprendizajes a nivel de asignatura y la calidad educativa a nivel de programa. • Para desarrollar herramientas útiles y fácil de usar y fáciles de usar, se requiere adoptar enfoques de investigación basada en diseño.

Estudio 3. Hilliger, I., Celis, S. & Pérez-Sanagustín, M. (2022). Engaged Versus Disengaged Teaching Staff: A Case Study of Continuous Curriculum Improvement in Higher Education. Higher Education Policy, 35, 81–101. https://doi.org/10.1057/s41307-020-00196-9	• El equipo docente se compromete con tareas de mejora continua cuando la recopilación de datos se centra en mejorar la enseñanza y el aprendizaje en diferentes niveles. • El equipo docente se compromete con tareas de mejora continua cuando recibe apoyo para sus prácticas docentes. • El equipo docente se desvincula de las tareas de mejora continua cuando las políticas institucionales están desalineadas, lo que les obliga a responder a lógicas contradictorias. • El equipo docente se desvincula de las tareas de mejora continua cuando su principal objetivo es dar respuesta a demandas externas.

La Tabla 1 resume los hallazgos obtenidos a partir de cada uno de los estudios previamente descritos. A partir del primer estudio, se identifican necesidades de herramientas analíticas en la región para distintos grupos de interés: las y los estudiantes necesitan retroalimentación de calidad y apoyo oportuno para mejorar su rendimiento académico, el equipo docente necesita alertas tempranas para ayudar al estudiantado inscritos en sus clases e información significativa para retroalimentar de la calidad de su enseñanza, y gestores necesitan información para evaluar si las intervenciones de apoyo implementadas permiten mejorar resultados a nivel institucional. A partir del segundo estudio, se establece que las herramientas de CA podrían ser útiles para recopilar una mayor variedad de evidencia respecto del logro de competencias del estudiantado a lo largo del plan de estudios, facilitando el monitoreo de los aprendizajes y el aseguramiento de la calidad. Finalmente, a partir del tercer estudio se confirma el compromiso del equipo docente con el aprendizaje de su estudiantado, lo cual se traduce en el interés por utilizar diversos métodos de evaluación para medir el logro de los resultados de los estudiantes, para luego utilizar esta evidencia para reflexionar sobre la efectividad de sus prácticas de enseñanza. Esta motivación intrínseca se ve particularmente reforzada cuando las y los profesores reciben apoyo para mejorar sus prácticas docentes.

Indistinto de la contribución de cada uno de los estudios previamente presentados, los contextos educativos son muy variables, por lo que es difícil generalizar los hallazgos obtenidos en campos como LA y CA. Sin embargo, al integrar los resultados de los tres estudios, se pueden extrapolar dos tipos de estrategias para favorecer la adopción de herramientas analíticas para la evaluación educativa y la mejora continua curricular.

La primera forma es estrategia es utilizar los resultados previamente descritos como un punto de partida para instalar procesos de mejora continua curricular. Esto implica realizar las siguientes tareas:

- **Con respecto a las necesidades de diferentes grupos de interés:**
  - Identificar los procesos y tecnologías existentes a nivel institucional que facilitan la entrega de retroalimentación oportuna a las y los estudiantes.
  - Identificar indicadores relevantes para evaluar la necesidad de intervenciones de apoyo y la calidad de las prácticas docentes.
- **Con respecto a las herramientas de CA:**
  - Identificar un proceso institucional que podría beneficiarse del uso de una herramienta analítica (por ejemplo, evaluación de la docencia del profesorado, recopilación de datos para la acreditación de programas, entrega de reconocimientos por excelencia docente).
  - Explorar cómo se recopilan y analizan los datos, con el fin de utilizarlos para informar las estrategias de renovación del plan de estudios.
  - Involucrar a gestores y equipo docente a lo largo de las fases de desarrollo de herramientas analíticas, de manera de abordar oportunamente problemas de usabilidad y funcionalidad.
- **Con respecto al compromiso del profesorado con la mejora continua:**
  - Aumentar las oportunidades para reflexionar sobre el proceso de enseñanza y aprendizaje (por ejemplo, involucrar a representantes estudiantiles en comités de programa para favorecer el intercambio de experiencias sobre elementos curriculares con el equipo docente).
  - Ofrecer apoyo al profesorado para alinear los métodos de enseñanza y las estrategias de evaluación utilizados en sus cursos con los resultados del aprendizaje esperados (por ejemplo, proporcionar al equipo docente de talleres al momento de planificar los cursos).
  - Alinear las políticas institucionales para reconocer el compromiso del equipo docente con tareas de mejora continua (dar cuenta de su participación en estrategias de renovación del plan de estudios o de sus asignaturas en su proceso de calificación).

La segunda estrategia consiste en utilizar los resultados previamente descritos como un enfoque metodológico, utilizando las metodologías de los estudios presentados para el diseño e implementación de herramientas analíticas para un entorno universitario específico. Este enfoque ligado al contexto implica la realización de las siguientes tareas:

- **Con respecto a las necesidades de las partes interesadas:**
  - Realizar entrevistas individuales semiestructuradas y grupos focales con diferentes partes interesadas de la educación superior para recopilar información sobre el uso actual y esperado de las herramientas analíticas.
  - Aplicar cuestionarios a estudiantes y miembros de la facultad para recopilar información sobre las expectativas de estas partes interesadas con respecto a la adopción de servicios de análisis de aprendizaje a nivel institucional.

- **Con respecto al diseño de herramientas basadas en CA:**
  - Adoptar un enfoque de investigación basado en el diseño para desarrollar e implementar una herramienta de CA a lo largo de un proceso iterativo:
    - Identificar requerimientos a partir de procesos existentes para la mejora continua curricular.
    - Evaluar el uso y la utilidad percibida de una primera versión de la implementación de una herramienta de CA.
    - Realizar un estudio de campo para anticipar el impacto de una herramienta de CA desde la perspectiva de diferentes usuarios.
- **Con respecto al compromiso del equipo docente:**
  - Analizar la evidencia documental recopilada de las tareas de evaluación de competencias y las reuniones del programa.
  - Realizar entrevistas semiestructuradas con directores de carrera y representantes del profesorado.

CONCLUSIONES

Este capítulo describe la evaluación educativa como un proceso permanente. No solo requiere la recolección y análisis de información, sino que también provee de datos a distintos grupos de interés. En primer lugar, le provee de información al estudiantado respecto a sus brechas de aprendizaje, de manera de que puedan realizar acciones para adquirir los conocimientos necesarios y nivelar su logro de habilidades. En segundo lugar, le provee de información al equipo docente para reflexionar sobre su práctica docente, y consecuentemente implementar ajustes curriculares en sus asignaturas. Y, en tercer lugar, le provee de información a gestores de los programas educativos para monitorear el logro de aprendizajes a lo largo de un plan de estudio, de manera de que puedan realizar intervenciones que favorezcan el éxito académico y el egreso oportuno de sus estudiantes.

En la actualidad, la recolección de información no es necesariamente una barrera para realizar una evaluación educativa más significativa y comprensiva del proceso de enseñanza y aprendizaje. Por lo contrario, las instituciones educativas están continuamente acumulando datos que se generan a partir de la interacción del estudiantado con distintos sistemas, tales como los LMS u otros servicios propios de cada institución. Esta acumulación se ha visto acelerada durante el contexto de la pandemia producto de la incorporación de tecnologías educativas y el uso de ambientes virtuales. Esto ha conllevado a que gestores y docentes puedan monitorear la participación del estudiantado en actividades de enseñanza y aprendizaje, e incluso conocer los resultados alcanzados en dichas actividades con una mayor celeridad. He ahí el compromiso requerido por parte de todos los actores educativos de sacar provecho a estos datos en favor de los resultados de aprendizaje.

Para facilitar la recolección, análisis, y visualización oportuna de datos educativos, investigadores han propuesto diferentes líneas de trabajo, incluyendo las analíticas de aprendizaje (LA) y la analítica curricular (CA). En la actualidad, diferentes estudios documentan

como estas líneas de trabajo han contribuido a la evaluación educativa y a la mejora continua curricular. Según sus hallazgos, existen distintas técnicas y herramientas que actores educativos podrían utilizar para realizar acciones oportunas que conlleven a una mejor calidad educativa. En este tenor, este capítulo presenta una sección que específicamente describe distintas aplicaciones en esta línea, dando cuenta del uso de técnicas estadísticas –tanto descriptivas como predictivas– de manera de facilitar el monitoreo de la participación del estudiantado, la entrega de retroalimentación, y la reflexión docente con respecto a distintos elementos curriculares.

Sin embargo, la mayoría de los trabajos documentados sobre LA y CA han sido generado en países anglosajones, por lo que no necesariamente tienen validez en otros contextos educativos. Para crear una mayor capacidad crítica en Latinoamérica u otras regiones, es clave instalar capacidades para el diseño e implementación de este tipo de estrategias y evaluar su contribución en diferentes contextos educativos. A la fecha, unos de los mayores esfuerzos que se ha realizado al respecto ha sido el proyecto LALA, generándose investigación y herramientas en distintas universidades de la región. Para diseminar las lecciones aprendidas a partir de este esfuerzo, este capítulo comparte los principales resultados de tres estudios que promueven la mejora continua curricular utilizando herramientas analíticas. A partir de la descripción de dichos estudios, no solo se espera promover la incorporación de herramientas analíticas para favorecer estrategias de evaluación en la sala de clases, sino que el uso de dichos resultados en instancias de discusión a nivel de programa educativo.

En ese sentido, una de las mayores contribuciones de este capítulo es resaltar la interdependencia entre el potencial de las herramientas analíticas y distintos elementos que son parte del proceso educativo. Por interdependencia, hacemos referencia a la mutua colaboración entre tareas, en favor de los aprendizajes. Es más, para sacar el mejor provecho posible a las herramientas analíticas, debe existir una alineación entre metodologías de enseñanza y estrategias de evaluación, y entre los resultados de aprendizaje evaluados a nivel de asignatura y las competencias definidas en el perfil de egreso. También debe existir una cultura de calidad compartida, la cual se manifieste en instancias de reflexión a nivel de programa y a nivel institucional. En ese sentido, las consideraciones presentadas en este capítulo pueden ser un buen punto de partida para iniciar procesos de mejora continua curricular en base a herramientas analíticas.

REFERENCIAS

Archer, E., & Prinsloo, P. (2020). Speaking the unspoken in learning analytics: troubling the defaults. *Assessment and Evaluation in Higher Education, 45*(6), 888–900. https://doi.org/10.1080/02602938.2019.1694863

Brown, M., DeMonbrun, R. M., & Teasley, S. (2018). Taken Together: Conceptualizing Students' Concurrent Course Enrollment across the Post- Secondary Curriculum Using Temporal Analytics. *Journal of Learning Analytics, 5*(3), 60–72. https://doi.org/10.18608/jla.2018.53.5

Cechinel, C., Ochoa, X., Lemos dos Santos, H., Carvalho Nunes, J. B., Rodés, V., & Marques Queiroga, E. (2020). Mapping Learning Analytics initiatives in Latin America. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.12941

Chou, C. Y., Tseng, S. F., Chih, W. C., Chen, Z. H., Chao, P. Y., Lai, K. R., Chan, C. L., Yu, L. C., & Lin, Y. L. (2015). Open Student Models of Core Competencies at the Curriculum Level: Using Learning Analytics for Student Reflection. *IEEE Transactions on Emerging Topics in Computing, 5*(1), 32–44. https://doi.org/10.1109/TETC.2015.2501805

Cobo, C., & Aguerrebere, C. (2018). Building capacity for learning analytics in Latin America. In C. Ping Lim & V. L. Tinio (Eds.), *Learning Analytics for the Global South* (Issue Learning Analitycs, pp. 63–67). Foundation for Information Technology Education and Development, Inc.

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research.* (P. E. Inc., Ed.), Educational Research (Fourth Edi, Vol. 4). Boston, Massachussetts. https://doi.org/10.1017/CBO9781107415324.004

Dicker, R., Garcia, M., Kelly, A., & Mulrooney, H. (2019). What does 'quality' in higher education mean? Perceptions of staff, students and employers. *Studies in Higher Education, 44*(8), 1425–1441. https://doi.org/10.1080/03075079.2018.1445987

Fernández-Marcha, A. (2009). *LA EVALUACIÓN DE LOS APRENDIZAJES EN LA UNIVERSIDAD: NUEVOS ENFOQUES*. Universidad Politécnica de Valencia. https://web.ua.es/es/ice/documentos/recursos/materiales/ev-aprendizajes.pdf

Greer, J., Molinaro, M., Ochoa, X., & Mckay, T. (2016). Learning Analytics for Curriculum and Program Quality Improvement (PCLA 2016). *Proceedings of the 6th International Conference on Learning Analytics & Knowledge* - LAK '16, 494–495. https://doi.org/10.1145/2883851.2883899

Hilliger, I., Aguirre, C., Miranda, C., Celis, S., & Pérez-Sanagustín, M. (2020). Design of a curriculum analytics tool to support continuous improvement processes in higher education. *Proceedings of the 10th International Conference on Learning Analytics & Knowledge* - LAK '20, 181–186. https://doi.org/10.1145/3375462.3375489

Hilliger, I., Celis, S., & Pérez-Sanagustín, M. (2020). Engaged Versus Disengaged Teaching Staff: A Case Study of Continuous Curriculum Improvement in Higher Education. *Higher Education Policy*, 1-21.

Hilliger, I., Celis, S. & Pérez-Sanagustín, M. (2022). Engaged Versus Disengaged Teaching Staff: A Case Study of Continuous Curriculum Improvement in Higher Education. Higher Education Policy, 35, 81–101. https://doi.org/10.1057/s41307-020-00196-9

Hilliger, I., Pérez-Sanagustín, M., Pérez-Álvarez, R., Henríquez, V., Guerra, J., Zuñiga-Prieto, M. Á., ... & De Laet, T. (2020). Leadership and Maturity: How Do They Affect Learning Analytics Adoption in Latin America? In *Adoption of Data Analytics in Higher Education Learning and Teaching* (pp. 305-326). Springer, Cham.

Hilliger, I., Aguirre, C., Miranda, C., Celis, S., & Pérez-Sanagustín, M. (2022). Lessons learned from designing a curriculum analytics tool for improving student learning and program quality. *Journal of computing in higher education,* 1-25. https://doi.org/10.1007/s12528-022-09315-4

IAU. (2020). *The Impact of Covid-19 on Higher Education around the World.* https://www.iau-aiu.net/IMG/pdf/iau_covid19_and_he_survey_report_final_may_2020.pdf

Ifenthaler, D., & Yau, J. Y.-K. (2020). Reflections on Different Learning Analytics Indicators for Supporting Study Success. *International Journal of Learning Analytics and Artificial Intelligence for Education (IJAI), 2*(2), 4. https://doi.org/10.3991/ijai.v2i2.15639

Maldonado-Mahauad, J., Hilliger, I., De Laet, T., Millecamp, M., Verbert, K., Ochoa, X., & Pérez-Sanagustín, M. (2018, March). The LALA project: Building capacity to use learning analytics to improve higher education in Latin America. In *companion proceedings of the 8th international learning analytics & knowledge conference* (pp. 630-637). Springer.

Kitto, K., Sarathy, N., Gromov, A., Liu, M., Musial, K., & Shum, S. B. (2020). Towards skills-based curriculum analytics: Can we automate the recognition of prior learning? *Proceedings of the 10th International Conference on Learning Analytics & Knowledge*, 171–180. https://doi.org/10.1145/3375462.3375526

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10*(3), 1–21. https://doi.org/10.1002/widm.1355

Siemens, G., & Gasevic, D. (2012). Guest Editorial - Learning and Knowledge Analytics. *Education Technology & Society, 15*(3), 1–2.

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior, 89*(July), 98–110. https://doi.org/10.1016/j.chb.2018.07.027

Wong, B. T. M., & Li, K. C. (2020). A review of learning analytics intervention in higher education (2011–2018). *Journal of Computers in Education, 7*(1), 7–28. https://doi.org/10.1007/s40692-019-00143-7

Yin, R. K. (2014). *Case Study Research Design and Methods* (5th ed.), Thousand Oaks, CA: Sage Publications.

Capítulo 45

LA INVESTIGACIÓN EN EVALUACIÓN EDUCATIVA

Mario Rueda Beltrán, Edna Luna Serrano y María del Pilar Rico Sánchez

"Si no sabes dónde vas, no importa el camino que elijas."
Alicia en el País de las Maravillas, *Lewis Carroll*

INTRODUCCIÓN

El propósito principal de este capítulo es compartir con los docentes universitarios algunos aspectos conceptuales, metodológicos y prácticos sobre la investigación en evaluación educativa, con la intención de que puedan incorporarse activamente a esta actividad.

En la primera parte se describen brevemente las características y los orígenes de la investigación educativa como antecedente de la investigación evaluativa. La coexistencia de distintos paradigmas en la investigación educativa ha hecho que una de sus características definitorias sea la presencia y uso cotidiano de diversas formas del quehacer científico que conviven en la tarea habitual de la producción de conocimientos, situación que ha trascendido a la investigación evaluativa.

En un segundo apartado se refiere la investigación evaluativa como una disciplina cuyo interés principal está centrado en el estudio de múltiples objetos con la intención de construir una valoración sobre ellos, con distintos énfasis en cuanto sus finalidades. Esta disciplina supone un trabajo riguroso en metodología, con respaldo de aspectos teóricos de alguna teoría específica o la aplicación de una metateoría. Tanto en el ámbito nacional como internacional, se reconoce que la evaluación se realiza en respuesta a las demandas sociales hechas por diferentes audiencias con necesidades heterogéneas de evaluación.

En un tercer apartado se delinea la investigación evaluativa derivada del desarrollo de la investigación educativa. En ésta se observa el predominio de los paradigmas empleados en la investigación educativa y las mismas características de las condiciones de producción. La situación para garantizar su avance es que toda producción de conocimiento dé cuenta de su proceso de gestación y desarrollo, de sus aciertos y fallas.

En una cuarta sección se revisa un breve panorama internacional y nacional de la investigación evaluativa. La referencia que se tomó para México fueron los estados de conocimiento elaborados por el Consejo Mexicano de Investigación Educativa, en la consideración del

análisis hecho a la producción académica de los pasados treinta años. Se plantea la investigación evaluativa como herramienta para los docentes, estudiantes y especialistas para intervenir en procesos que conduzcan a la comprensión de los fenómenos educativos y la transformación de la evaluación.

En la quinta sección se reconoce, en el ámbito nacional, la existencia de una pluralidad de objetos de evaluación –constituidos en líneas de investigación y se delinea una posible agenda por desahogar–. Se hace un llamado a desplegar trabajos de evaluación científica, en dónde se examinen las ventajas y limitaciones de cada aproximación metodológica, se empleen diversas estrategias e instrumentos de obtención de información y sea tomado en cuenta el contexto. El desarrollo diferencial de los temas abre la posibilidad de incorporarse activamente en la producción del conocimiento y surge como indispensable la incorporación de más agentes que contribuyan a consolidar el crecimiento de la investigación evaluativa. A la par, se sugieren algunas pistas para la incorporación de nuevos investigadores que consoliden y amplíen el campo de la investigación evaluativa.

LA INVESTIGACIÓN EDUCATIVA

La investigación educativa es una actividad relativamente joven, surgida de la conjunción de diversas disciplinas interesadas en el estudio del campo de la educación. En sus orígenes, surgió como parte del desarrollo de algunas disciplinas como la antropología, la psicología y la sociología, interesadas por estudiar algunos de los múltiples aspectos de la educación. La psicología educativa, por ejemplo, es una rama derivada de la psicología general que ha influido considerablemente en muchas de las prácticas iniciales y presentes en la investigación educativa, como es el caso de los estudios sobre el aprendizaje y de la evaluación, entre otros temas derivados de la compleja experiencia escolar.

La investigación educativa, ahora, goza de reconocimiento como una actividad profesional, después de un amplio recorrido previo a su constitución, derivado del desarrollo de las disciplinas que conformaron ramas de conocimiento especializadas. Disciplinas que una vez reagrupadas en un nuevo campo profesional, adquirieron una naciente identidad, con las dificultades que entraña la convivencia permanente y prolongada con múltiples tradiciones para producir conocimientos, desarrollar teorías, emplear metodologías e inclusive la promoción de preferencias temáticas originarias de cada una de las disciplinas de procedencia.

Se podría afirmar que, en el caso de la investigación educativa, no es posible identificar un solo paradigma dominante a lo largo de toda su constitución y desarrollo, seguido del surgimiento posterior de una crítica sistemática que diera paso a la formulación de uno nuevo en sustitución del anterior, como ha ocurrido en otros campos científicos. Aquí, haciendo alusión de forma muy general a paradigma, como el conjunto de creencias, valores, técnicas y preferencias por el estudio de ciertos temas, compartidos por los miembros de una comunidad científica en particular (Kuhn, 2013), por lo que se sostiene que, en la investigación educativa, desde su origen, han estado presentes de forma simultánea distintos paradigmas.

La coexistencia de los paradigmas en la investigación educativa ha hecho que una de sus características definitorias sea la presencia y uso cotidiano de diferentes formas del quehacer científico que conviven en la tarea habitual de la producción de conocimientos. Lo que conlleva a un desdibujamiento de la labor disciplinaria de quienes han construido o contribuyen a nutrir este nuevo campo. Así, es frecuente que un psicólogo emplee tradiciones sociológicas o antropológicas en sus indagaciones sobre temas de la educación, o un sociólogo recurra a metodologías más tradicionalmente identificadas con la actividad de los psicólogos. De ahí la gran dificultad, en parte, de la comunicación entre colegas en espacios compartidos, como pueden ser la creación y funcionamiento de una revista especializada, o un organismo de gestión de financiamiento para el desarrollo de proyectos, que pretendan contribuir al esclarecimiento de los estándares de calidad de esta actividad, o bien, para el fortalecimiento mismo de la investigación educativa como disciplina profesional. Lejos de constituir un problema la convivencia de múltiples paradigmas, de hecho, es una oportunidad para quienes incursionan en la investigación, adherirse al paradigma más apropiado a las características del objeto de estudio de interés, e inclusive hacer uso de alguno de su preferencia, para adscribirse de forma natural a sus reglas de operación que le permitirán aportar al conocimiento sistemático del objeto estudiado.

En algún momento del desarrollo de la investigación educativa, el conocimiento reconocido con estatus científico era aquel identificado con características como la objetividad, la medición, los diseños experimentales, la identificación de variables, la replicabilidad, entre otras. Por lo que no resultó extraño que, durante una etapa, los métodos cuantitativos, con sus procesos y herramientas, gozaran de una mayor aceptación en la comunidad científica, mientras que los métodos cualitativos tuvieron que desarrollar múltiples estrategias para legitimar sus aportaciones al conocimiento de los fenómenos propios del campo de la educación. Después de varias décadas de discusión sobre las bondades y límites de ambas metodologías, en la actualidad la tendencia dominante se orienta a reconocer las características propias de cada una y a realizar esfuerzos por lograr su complementariedad en el estudio de los fenómenos educativos. Siguen en construcción permanente las respuestas al reto de establecer criterios de cientificidad, acordados colectivamente, en comunidades conformadas por profesionales con formación en tradiciones paradigmáticas diversas.

En México, si se parte del primer diagnóstico nacional de la investigación educativa, realizado bajo el patrocinio del Consejo Nacional de Ciencia y Tecnología (Conacyt) en 1981, se puede afirmar que después de cuatro décadas de esfuerzos individuales y colectivos se constatan indicadores que dan testimonio de la configuración de este relativamente joven quehacer profesional, denominado investigación educativa. El número de profesionales dedicados principalmente a esta actividad ha crecido, así como los establecimientos especializados en dónde se practica la investigación educativa de forma casi exclusiva, también es notoria la presencia de más programas de formación profesional con claro énfasis en la investigación y se cuenta con una mayor variedad de foros académicos y publicaciones especializadas; a la par, han crecido distintas redes de académicos que se agrupan y cultivan temas particulares de investigación. Sin duda, una organización como el Consejo Mexicano de Investigación

Educativa (COMIE), fundado en 1993, ha contribuido de forma significativa a la consolidación de este campo profesional a través de múltiples iniciativas, dentro de las que destacan particularmente, la organización bianual de Congresos Nacionales de Investigación Educativa, la labor ininterrumpida de la Revista Mexicana de Investigación Educativa, la elaboración de los estados de conocimiento de la producción en investigación por cada década en el país y muchas otras acciones de aliento a esta actividad con repercusiones en todo el territorio nacional.

Como parte del surgimiento, desarrollo y consolidación de la investigación educativa se observa un panorama de estudios sobre muy diversos temas del campo de la educación, y la presencia de metodologías de corte cualitativo, cuantitativo o mixto; las diferentes colecciones de libros sobre los estados del conocimiento dan cuenta de la pluralidad de líneas de investigación y temáticas desarrolladas en México[1].

LA INVESTIGACIÓN EVALUATIVA

En el conglomerado de áreas atendidas por la investigación educativa se encuentran los trabajos sobre evaluación, cuyo interés principal está orientado al estudio de múltiples objetos, como pueden ser las políticas, el sistema escolar, el aprendizaje de los estudiantes, las actividades de los profesores y personal directivo, la efectividad de un programa, entre otros. Como punto central, comparten la intención de construir una valoración sobre el objeto en cuestión, con distintos énfasis en cuanto sus finalidades. En algunos casos el propósito principal es el mejoramiento del objeto evaluado, en otros, la constatación del cumplimiento de criterios o estándares previamente definidos, o el conocimiento de las perspectivas de los diferentes actores involucrados en la experiencia estudiada, o la obtención de información para tomar decisiones sobre la continuidad, modificación o eliminación de un programa. Ha sido tal el interés creciente y la diversidad de los objetos de evaluación que, a este agrupamiento de trabajos actualmente se le reconoce como investigación evaluativa.

La investigación evaluativa es una rama derivada de la investigación educativa, y como tal, ha heredado las influencias y características de este campo general. En particular, se puede identificar el predominio de los paradigmas empleados por la investigación educativa, así como advertir las mismas características de las condiciones de producción. Asimismo, ambas actividades comparten el desafío de enfrentar retos particulares, como descifrar la naturaleza misma del objeto estudiado, el tratamiento de la complejidad de los temas abordados, y las motivaciones iniciales del investigador. Motivaciones surgidas del cultivo de un interés individual o derivado de un encargo institucional, una iniciativa desplegada de las prácticas profesionales en el campo de la educación, o bien, de las tradiciones disciplinarias, o una combinación de todas ellas.

En este sentido, de forma generalizada se asume que la evaluación educativa es una modalidad de investigación aplicada en educación (Jornet, Perales y González-Such, 2020;

[1] http://www.comie.org.mx/v5/sitio/estados-de-conocimiento/

Escudero, 2016), desarrollada en el marco de las ciencias sociales y las humanidades (Scriven, 2003). Sin embargo, cabe la aclaración, no todo trabajo de evaluación cumple con los requerimientos para ser considerado como una actividad de investigación. Al respecto, los argumentos planteados por Scriven y Stake clarifican las condiciones necesarias que deberían cumplir los estudios para integrarse al campo de conocimientos reconocidos como elementos propios de la investigación evaluativa.

La investigación evaluativa de acuerdo con Scriven (1991b), identifica el trabajo realizado a través de la aplicación rutinaria de principios y técnicas validados. Reconoce que ésta tiene subáreas que se corresponden en función de los diversos campos de aplicación de la evaluación, donde se realiza tanto investigación teórica como aplicada; así como la naturaleza del objeto central de la evaluación con sus respectivas subdivisiones (las políticas, los estudiantes, los docentes, los programas y las organizaciones, como los sistemas educativos o las instituciones). Desde esta perspectiva, para Scriven (1991b) la investigación en evaluación siempre remite de forma explícita a un trabajo riguroso en metodología, al respaldo otorgado por los aspectos teóricos de alguna teoría específica o la aplicación de una metateoría.

Las anteriores nociones se esclarecen con los planteamientos adicionales de Stake (2006), quien conceptualizó a la investigación evaluativa como el estudio de la calidad con un afán paralelo de generalización. En sus propias palabras, "todos los estudios de evaluación formal son investigación, pero solo los que aspiran a la generalización de algo a una población de casos se conocen con el nombre específico de investigación evaluativa" (p. 90).

Las características que definen a la investigación evaluativa se derivan del quehacer propio del evaluador que se realiza bajo condiciones y en problemas que difieren de otras actividades de investigación (McLaughlin y Phillips, 1991). En México como en otros países, la evaluación se realiza en respuesta a las demandas sociales hechas por diferentes audiencias con necesidades heterogéneas de evaluación (Luna y Torquemada, 2008). Lo que remite a considerar que una definición extendida de evaluación la define como el proceso de determinar el mérito, valor o significancia de algo (Scriven, 1991b). En consecuencia, la investigación evaluativa en el marco de los criterios de calidad del paradigma utilizado conduce a una opinión fundamentada de que algo es de cierto modo, lo que guía a la construcción de un juicio acerca de su valor.

A lo largo del desarrollo de la investigación evaluativa, ha habido intensos debates sobre cuál sería la mejor manera de definirla y aunque la polémica continúa abierta, esto no ha obstaculizado el devenir de su crecimiento. Además de las posturas que optan por poner el énfasis de la definición de evaluación, en el logro de los objetivos educativos, o la determinación del mérito o valía, otras proponen como punto clave la obtención de información para tomar decisiones, o bien refieren a una actividad que comprende tanto la descripción como el juicio, o a una investigación sistemática del valor o mérito de algún objeto (House, 1986).

Además de la discusión sobre las distintas definiciones de la investigación evaluativa, algunas voces plantean la conveniencia de adicionalmente formularse preguntas que pudieran dilucidar aspectos fundamentales de los procesos de evaluación. Como aquellas, que contribuyan a esclarecer si la evaluación se ubica como proactiva, destinada a servir para

tomar decisiones, o más bien se trata de la evaluación retroactiva para auxiliar a la rendición de cuentas. Así como, la clarificación para conocer si el proceso de evaluación abonará a perfilar una función psicológica o sociopolítica, o sencillamente si se está frente a una función del ejercicio de la autoridad (Phillips, 2018).

El panorama de los productos de la investigación evaluativa muestra que estos se desarrollan bajo el sustento de los paradigmas dominantes. Por un lado, el paradigma cuantitativo, importado de las ciencias naturales y exactas, positivista, tendiente a la explicación de los fenómenos; y el paradigma cualitativo, generado por las ciencias humanas, subjetivista, dirigido a la comprensión. Dado el origen epistemológico de cada uno, parten de supuestos distintos, por lo que generan metodologías de investigación diferentes. En el primero, se originan metodologías cuantitativas o experimentales, hipotético deductivas; y en el segundo cualitativas o comprensivas (interpretativas), empírico-inductivas (Goetz y LeCompte, 1988).

Cronbach (1982), propuso la utilización de métodos combinados para evaluar de manera más eficiente los objetos de interés. El enfoque de los métodos combinados plantea que tanto las técnicas como los análisis no son privativos de una metodología en particular (Bryman, 1984) y deja al investigador, en el contexto de un estudio específico, la responsabilidad de fundamentar las combinaciones que va a utilizar. Esta tendencia, en respuesta a las posturas que promueven el conflicto entre lo cuantitativo y lo cualitativo, sostiene la perspectiva de la complementariedad de los métodos más que de su oposición. En la actualidad, se puede sostener que la postura asumida por el investigador en torno a cómo concibe el objeto de estudio, determina en gran medida el paradigma que mejor puede sustentar argumentaciones de calidad: confiables, válidas y representativas.

Una discusión inherente tanto en el ámbito de la investigación evaluativa como en la práctica de la evaluación alude a sus dos funciones críticas ampliamente reconocidas. La función *formativa* y la función *sumativa*. La función *formativa* es una evaluación diseñada, realizada y utilizada para el mejoramiento de lo evaluado; y la función *sumativa* diseñada y usada con el propósito de juzgar el mérito o valor del objeto de evaluación (Scriven, 1991b; Scriven, 1996). Los términos inicialmente planteados por Scriven en 1963 representan una clasificación particular de las funciones de la evaluación, a partir de la cual el campo de la evaluación educativa ha evolucionado.

Esta aparentemente sencilla clasificación ha tenido un gran impacto en las comunidades académicas y ha suscitado múltiples debates por lo que cabe tener presentes algunas precisiones formuladas por el mismo autor. Para él (1991a; 1996) el elemento que distingue a una y otra es *el propósito*, la intención original de la evaluación. La distinción no es intrínseca al procedimiento, depende del contexto. La evaluación formativa desde su origen se realiza sobre la base del mejoramiento mientras que la sumativa pone el acento de su orientación en la emisión de un juicio de valor. Así, los dos tipos de evaluaciones pueden considerarse como igualmente útiles, y dependerá de las circunstancias que se asuma alguna de ellas.

Debido al momento en que se encuentra el desarrollo de la investigación educativa, se puede afirmar que aún sigue sin resolverse la falta de consenso en la comunidad científica respecto de la conceptualización misma de qué es investigar y cuáles debieran ser sus

"productos" y sus correspondientes criterios de aceptabilidad y reconocimiento (Rodríguez, 2005). Situación también compartida por la investigación evaluativa, por lo que es necesario que quienes la realicen, abonen a la consolidación de un ambiente que promueva la explicitación de las condiciones que originaron el proyecto, las razones de adscripción a alguna tradición investigativa, las preferencias teóricas y metodológicas empleadas, así como las características de los medios de comunicación de sus hallazgos. Es decir, se debe propugnar para que toda producción de conocimiento dé cuenta de su proceso de gestación y desarrollo, de sus aciertos y fallas, y de las formas en que se lograron, para poder apreciar en su justa dimensión, los resultados de esta actividad. Con el cumplimiento de estas directrices, entre otros posibles recursos, se podrá avanzar en la consolidación del campo de la investigación evaluativa.

PANORAMA INTERNACIONAL Y NACIONAL DE LA INVESTIGACIÓN EVALUATIVA

Durante la década de 1960, la investigación evaluativa surge en Estados Unidos y en Europa, particularmente en Inglaterra como un campo de estudio relevante que logró su institucionalización. Los primeros trabajos enfocaron su interés en ámbitos referentes a la salud y la educación. Actualmente, es notable el desarrollo de la evaluación y de la investigación evaluativa en Canadá, Australia, Israel, África y países de América Latina (Worthen, Sanders y Fitzpatrick, 1997). Por lo tanto, se ha observado un interés y estudio creciente del tema, lo que ha generado el desarrollo de asociaciones de profesionales en investigación en evaluación y el surgimiento de revistas especializadas. Entre algunas de las asociaciones, se puede mencionar a la *American Evaluation Association*, la *Canadian Evaluation Society,* la *European Evaluation Society*, la *Australian Evaluation Society*, la *African Evaluation Association*, la *United Kingdom Evaluation Societ*y, y la *Associazione Italiana de Valutazione*. Por su parte, algunos ejemplos de revistas especializadas los ubicamos en: *The American Journal of Evaluation*, *New Directions for Evaluation*, *Educational Evaluation and Policy Analysis*, *Canadian Journal of Program Evaluation*, *Revista Iberoamericana de Evaluación Educativa* y *RELIEVE*. A través de estos recursos se han realizado diversos encuentros y foros de discusión para el intercambio de experiencias, planteamiento de problemas, métodos y modelos de evaluación, que han favorecido el desarrollo de este campo de conocimiento.

En el caso de México, la evolución de la investigación evaluativa se puede reconstruir a través del análisis de los textos elaborados para resguardar la memoria del quehacer profesional de los investigadores educativos en el país, concebidos y organizados por el COMIE. En esta colección de libros sobre los estados del conocimiento de la investigación educativa nacional, elaborados cada diez años, se da cuenta del surgimiento de temas, de redes nacionales de investigadores, de asociaciones e instituciones interesadas en la investigación de muy diversos tópicos, dentro de los que ahora se destaca solamente lo referido a la evaluación.

En el primer Estado del Conocimiento "La Investigación Educativa en los Ochenta, Perspectiva para los Noventa", se recopiló la investigación educativa mexicana realizada de

1982 a 1992, en él se desarrolló el tema de la Evaluación de la Educación, y las categorías derivadas del análisis de los documentos publicados, reflejan las características globales de la producción; así se habla de las condiciones institucionales de realización de la investigación en evaluación educativa, sus problemas epistemológicos, la evaluación institucional, la evaluación curricular y la evaluación de los docentes. Tanto por la naturaleza y el número de categorías empleadas para agrupar los trabajos, se puede interpretar una etapa incipiente de la investigación evaluativa con el abordaje de unos cuantos temas de interés. Además de que en el marco del conjunto de reportes revisados se puntualizó continuar con la indagación de los impactos de las políticas nacionales de evaluación, y con investigaciones que brinden un sustento para tomar decisiones, a partir de trabajos sistemáticos con marcos teóricos sólidos (Carrión, 1995). Por lo que se puede interpretar que predominó en el conjunto de trabajos de evaluación una perspectiva sumativa y de orientación para tomar decisiones.

En el siguiente Estado del Conocimiento que abarca la década de 1992 a 2002, el tema de la investigación evaluativa no se destacó de forma particular, los trabajos referidos a este campo se incluyeron como parte o secciones especiales de objetos de investigación más amplios como la planeación, el currículo o las políticas educativas. Este hecho puede interpretarse como que predominaba un interés menor entre los académicos hacia la evaluación, una presencia poco significativa de reportes sobre el tema, en general que, las actividades de investigación educativa aún no eran tan requeridas por las instituciones. Además, las políticas gubernamentales todavía no incluían a la evaluación como elemento esencial para la orientación y regulación del sistema educativo. Como ya es conocido, puede prevalecer que el abordaje de los temas de investigación responda al interés de los académicos por conocer los efectos que las políticas en turno provocan en el sistema. Siguiendo el hilo conductor del desarrollo de la investigación evaluativa se puede afirmar que, en esta década, según lo reportado en el Estado de Conocimiento, no hay un reconocimiento explícito de este campo como área independiente.

Es hasta el Estado del Conocimiento de la década de 2002 al 2011 que adquiere mayor peso la presencia de la investigación evaluativa, en el conjunto de los campos de estudio del ámbito educativo. Esta colección da cuenta del Estado del Conocimiento sobre la evaluación educativa en México en este periodo. El libro del campo temático "*Educación y ciencia: políticas y producción del conocimiento 2002-2011*", incluyó el capítulo "*La investigación sobre la evaluación educativa*", el cual mostró la conformación de un campo de conocimiento complejo y diverso, tanto en las temáticas abordadas como en los métodos utilizados. El conjunto de estudios fue tan abundante que su análisis condujo a su clasificación en categorías como la evaluación del aprendizaje, las tendencias teóricas y metodológicas; la evaluación de la docencia en el sistema educativo nacional; la evaluación de políticas educativas, el sistema educativo mexicano, las instituciones y los programas (Luna, 2013).

Después de casi 30 años respecto del primer Estado del Conocimiento (1982-1992), el campo de la evaluación educativa logró un mayor reconocimiento, dado que, por primera vez se le considera como un área independiente. Las líneas de investigación contempladas para el periodo 2012-2021 son en número y especificidad reflejo del desarrollo de la

investigación evaluativa. Tan solo la enumeración y el agrupamiento de lo publicado son elocuentes: la evaluación docente y desarrollo profesional, la evaluación de directivos; la evaluación del y para el aprendizaje; la evaluación institucional y de centros educativos; la evaluación curricular y acreditación de programas; la evaluación de académicos e investigadores en educación superior; y evaluación de políticas educativas. Este desglose es una muestra de la existencia tanto de la diversidad de grupos de investigación dedicados a la investigación evaluativa, como de la cantidad de investigaciones desarrolladas durante la década analizada.

La investigación evaluativa como campo de estudio convoca a docentes, estudiantes y especialistas a contribuir desde diversos ámbitos para intervenir en procesos que nos lleven a la comprensión, la transformación y a la reflexión en torno al ejercicio de la evaluación. Por lo tanto, es imprescindible mirar a la investigación evaluativa como un campo en constante evolución, como motor de cambio para lograr su consolidación, continuidad y profesionalización. La investigación evaluativa debe tener, entre otros propósitos, estimular la discusión sobre estos temas, proponer a investigadores y docentes una experiencia intelectual genuina. Una oportunidad para la reflexión y el análisis para generar preguntas que conduzcan a construir las respuestas para enfrentar los complejos retos educativos contemporáneos.

LA AGENDA NACIONAL PENDIENTE DE LA INVESTIGACIÓN EVALUATIVA

En la actualidad en México, la investigación evaluativa se ha conformado como un campo del conocimiento complejo y diverso. Se reconoce una multiplicidad de objetos de evaluación constituidos en líneas de investigación dentro del propio campo. La producción de trabajos de investigación muestra la relevancia del campo de estudio ya sea en evaluación del y para los aprendizajes, del profesorado, o de políticas y programas. De forma generalizada, el conocimiento acumulado pondera, sobre todo, el uso de los resultados de la evaluación para mejorar la calidad educativa, entendida como guiar el logro de los aprendizajes de los estudiantes, optimizar las prácticas pedagógicas del profesor y determinar la eficacia y pertinencia de programas y políticas.

Se ha demostrado la complejidad de los objetos de estudio, por lo tanto, se requiere actuar con cautela y se hace necesario emplear aproximaciones multidimensionales y métodos diversos. En este sentido, los trabajos reclaman una evaluación científica, que conozca a profundidad las ventajas y limitaciones de cada aproximación metodológica y que haga uso de diversas estrategias e instrumentos de obtención de información. Además, de que sea sensible al contexto, adecuada a las peculiaridades de los estudiantes, el profesorado o los programas educativos. El recuento de la producción refleja un desarrollo diferencial entre las líneas de investigación, un mayor progreso se ha mostrado en los temas de evaluación del y para los aprendizajes, la evaluación de la docencia y, la evaluación de políticas educativas del sistema, las instituciones y los programas (Luna, 2013). Al mismo tiempo, es observable que al interior de cada tema hay disparidad en el desarrollo de los objetos de investigación; a manera de ejemplo, la evaluación de la docencia en el nivel superior tiene una mayor

evolución que lo investigado en el nivel de preescolar. Reconocer este hecho abre la posibilidad de incorporarse activamente en la producción del conocimiento que nos permita socialmente enfrentar los desafíos propios de cada tema y sus vinculaciones con el conjunto de tópicos afines.

Si bien cada línea de investigación tiene una agenda propia, se coincide en la necesidad de profundizar en el conocimiento que ayude a realizar evaluaciones justas y pertinentes a cada persona u objeto de evaluación, así como establecer canales adecuados de comunicación que permitan persuadir a los encargados de tomar decisiones, hacerlo con el apoyo del conocimiento acumulado en la investigación evaluativa; con el fin de minimizar las discrepancias identificadas entre el conocimiento acumulado y las prácticas de evaluación, ampliamente documentado por las investigaciones realizadas.

En el marco de todo lo anteriormente expuesto, surge como necesaria la incorporación de más agentes que contribuyan a consolidar el crecimiento de la investigación evaluativa. La consolidación es posible en la medida que se cuenta con fuentes especializadas accesibles y la memoria del quehacer investigativo nacional de las últimas tres décadas en las que se consigna quienes han desarrollado trabajos de investigación, los temas, los aspectos teóricos y metodológicos empleados y los principales hallazgos disponibles.

Cada uno de los agentes, los políticos que diseñan orientaciones generales para el sistema educativo, los directivos que ponen en marcha los procesos de evaluación, los profesores que los conducen directamente, y los estudiantes que cada vez más amplían su participación en ellos, todos pueden contribuir a través de distintas iniciativas al crecimiento del campo de la investigación evaluativa. De acuerdo con lo observado en otras disciplinas académicas, es muy importante la sistematización del conocimiento que se produce al ofrecer testimonio y analizar críticamente las prácticas profesionales que se realizan, así como procurar el aliento de las discusiones teóricas y metodológicas que surjan en torno al tema de la evaluación, ya que la acumulación del conocimiento sistematizado es un elemento clave para garantizar entre todos los miembros de la comunidad académica la consolidación del campo. Así como establecer las condiciones para la incorporación de nuevos investigadores que aseguren la continuidad y crecimiento del interés sobre los diversos temas estudiados.

Una posible ruta para incorporarse como investigador evaluativo puede iniciar con la identificación de algún aspecto de la evaluación que atrape nuestro interés, y que nos pueda guiar para hacer una búsqueda que nos conduzca a conocer la mayor parte del conocimiento disponible sobre ese tema en particular. Algunos investigadores educativos se refieren a este vínculo emocional establecido con su tema elegido, expresando, por ejemplo, que “el juego me convirtió en investigador”. Esta expresión muestra que el interés por estudiar “el juego” impulsó al investigador a querer saber más sobre él. A la vez, su motivación por saber más le dio la oportunidad de iniciar a profundidad una indagación sobre el conocimiento disponible sobre el objeto de su interés. Así como descubrir las distintas estrategias empleadas para conformar el conjunto de conocimientos hasta ese momento disponible para todos. Tanto los distintos puntos de vista surgidos de los trabajos orientados por diferentes disciplinas, como las teorías y las metodologías que se han empleado para elaborar las contribuciones

al saber acumulado del objeto elegido. De tal manera que el estudioso pueda reconocer una variedad de distintos enfoques, recursos metodológicos e instrumentos que podrá ponderar para asumir con cuáles de esos elementos disponibles él mismo podría emprender una indagación propia, un proyecto de investigación, con el que pudiera también sumarse con contribuciones originales, al conocimiento acumulado y disponible.

El campo de la investigación evaluativa es fascinante y una herramienta poderosa para contribuir al conocimiento sistemático de los hechos educativos, comprender su naturaleza, explicar su actuación y ofrecer guías para su perfeccionamiento. ¡Bienvenidos al campo de la investigación evaluativa!

RECURSOS EN LÍNEA

Recurso
Consejo Mexicano de Investigación Educativa A.C. http://www.comie.org.mx
Red Iberoamericana de Investigadores sobre Evaluación de la Docencia RIIED http://riied.ens.uabc.mx/
Evaluación Educativa CUAIIED https://cuaieed.unam.mx/evaluacion_educativa
Consejo de Evaluación Educativa de la UNAM https://cuaieed.unam.mx/evaluacion_educativa#evaluacion-educativa
Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia (CUAIEED). https://cuaieed.unam.mx/
African Evaluation Association https://afrea.org/
American Evaluation Association https://www.eval.org/
Associazione Italiana de Valutazione https://valutazioneitaliana.eu/
American Journal of Evaluation https://journals.sagepub.com/home/aje
Australian Evaluation Society https://www.aes.asn.au/
Canadian Evaluation Society https://evaluationcanada.ca/

Canadian Journal of Program Evaluation https://journalhosting.ucalgary.ca/index.php/cjpe/index
Educational Evaluation and Policy Analysis https://journals.sagepub.com/home/epa
European Evaluation Society https://europeanevaluation.org/
Evaluation https://journals.sagepub.com/home/evi
New Directions For Evaluation https://onlinelibrary.wiley.com/journal/1534875x
Revista Iberoamericana de Evaluación Educativa https://revistas.uam.es/riee
UK Evaluation Society https://www.evaluation.org.uk/
Revista Electrónica de Investigación y Evaluación RELIEVE https://revistaseug.ugr.es/index.php/RELIEVE/index

REFERENCIAS

Bryman, A. (1984). The debate about quantitative and qualitative research: A question of method or epistemology? *The British Journal of Sociology*, *1*. 75-92.

Carrión, Carmen, (1995) *Estado del Conocimiento 17, Evaluación de la educación*. México: ANUIES. Recuperado de https://www.ses.unam.mx/curso2007/pdf/CARRION1995.pdf

Cronbach, L. J. (1982). *Designing Evaluations of Educational and Social Programs*. San Francisco, California: Jossey-Bass.

Escudero, T. (2016). La investigación evaluativa en el siglo XXI: Un instrumento para el desarrollo educativo y social cada vez más relevante. *RELIEVE*, *22*(1). doi: http://dx.doi.org/10.7203/relieve.22.1.8164

Goetz, J. P., y LeCompte, M. (1988). *Etnografía y diseño cualitativa en investigación cualitativa*. Madrid: Morata.

House, E. (1986). Evaluation and Legitimacy. En E. House (Ed.), *New directions in educational evaluation*. 15-29. Gran Bretaña: Falmer Press.

Jornet, J. M., Perales, M. J., y González-Such, J. (2020). El concepto de validez de los procesos de evaluación de la docencia. *Revista Española de Pedagogía, 78*(276), 233-252. doi: https://doi.org/10.22550/REP78-2-2020-01

Kuhn, T. (2013). *La estructura de las revoluciones científicas*. Ciudad de México: Fondo de Cultura Económica.

Luna, E, (2013). *La investigación sobre evaluación educativa.* En A. Maldonado (Coord.), *Educación y Ciencia Políticas y producción de conocimiento 2002-2011* (pp. 379-384). México: ANUIES-COMIE. Recuperado de http://www.comie.org.mx/v5/sitio/wp-content/uploads/2020/08/Educaci%C3%B3n-y-Ciencia.pdf

Luna, E., y Torquemada, A. (2008). Los cuestionarios de evaluación de la docencia por los alumnos: balance y perspectivas de su agenda. *Revista Electrónica de Investigación Educativa,* Especial. Recuperado de https://redie.uabc.mx/redie/article/view/201

McLaughlin, M., y Phillips, D. C. (1991). The formative/summative distinction. En M. McLaughlin y D. C. Phillips (Eds.). *Evaluation and Education: al Quarter Century* (pp. 18). Chicago: University of Chicago Press/NSSE.

Phillips, D. C. (2018). The many functions of evaluation in education. *Education Policy Analysis Archives, 26*(46). Recuperado de http://dx.doi.org/10.14507/epaa.26.3811

Rodríguez, P. G. (2005). *Linderos. Diálogos sobre investigación educativa*. México: COMIE, CEE, SEB, ITESO.

Scriven (1991a). *Evaluation Thesaurus* (4ta. ed.). Newbury Park, California: Sage publications.

Scriven (1991b). Beyond formative and summative evaluation. En M. McLaughlin y D. C. Phillips (Eds.). *Evaluation and Education: at Quarter Century* (pp. 19-64). Chicago: University of Chicago Press/NSSE.

Scriven, M. (1996). Types of Evaluation and Types of Evaluator. *Evaluation Practice*,*17*(2), 151-161.

Scriven, M. (2003). Evaluation theory and metatheory. En T. Kellaghan, D. L. Stufflebeam, L. R. Wingate (Eds.). *International Handbook of Educational Evaluation* (pp. 15-30). Estados Unidos: Springer.

Stake, R. (2006). *Evaluación comprensiva y evaluación basada en estándares.* Barcelona: Graó.

Worthen, B., Sanders, J., Fitzpatrick, J. (1997). *Program Evaluation. Alternative Approaches and Practical Guidelines* (2da. ed.). Estados Unidos: Longman Publishers.

GLOSARIO DE EVALUACIÓN EDUCATIVA

Nancy Sofía Contreras Michel, María Elena Pérez Rivera, Enrique Ricardo Buzo Casanova,
Elibidú Ortega Sánchez, Careli Johana Herrera Penilla, Adrián Martínez González

*"Toda palabra dice algo más de lo que debiera
y también menos de lo que debiera expresar."*
José Ortega y Gasset

INTRODUCCIÓN

Este glosario comprende conceptos relevantes en el ámbito de la evaluación educativa y se pone de manifiesto que en él existen diferentes perspectivas y matices sobre la evaluación, considerando que se trata de un libro que se destaca por la participación de un gran número de autores que se distinguen por generar conocimiento en el contexto actual de la educación.

Asimismo, la experticia con la que cuentan contribuye en gran manera al debate en torno a las prácticas de evaluación y el empleo de diferentes enfoques y métodos que enriquecen el proceso de enseñanza-aprendizaje-evaluación.

En gran medida el contenido de este glosario se elaboró con base en la experiencia en evaluación educativa con la que cuenta la CUAIEED en su práctica diaria, sin embargo, es importante difundir en los espacios de aprendizaje el diálogo interdisciplinario, el pensamiento crítico y la inclusión de diferentes perspectivas como agentes dinámicos en la construcción de la cultura de la evaluación.

Por lo anterior se invita a los usuarios a establecer contacto con nosotros, a fin de recibir sus valiosas aportaciones sobre los conceptos incluidos en este glosario, a fin de robustecerlo y que sirva como guía para los docentes que realizan actividades de evaluación del, para y como aprendizaje.

También les recomendamos revisar los glosarios de la UNESCO[1] y de la SEP[2] sobre este tema.

1 Instituto Internacional de Planeamiento de la Educación de la UNESCO. (s. f.). *Glosario*. Learning Portal. Recuperado el 26 de enero de 2022, de https://learningportal.iiep.unesco.org/es/glossary

2 Secretaría de Educación Pública. (s.f.). *Glosario de términos evaluativos*. Recuperado el 3 de febrero de 2022, de http://www.dgadae.sep.gob.mx/LnkDgep/DGADAE_Glosario_Terminos_Evaluativos.pdf

TÉRMINO	DEFINICIÓN
Abandono escolar	Situación en la que un alumno, habiendo estado inscrito en un ciclo escolar, no se reinscribe en ciclos subsecuentes.
Acreditación	Proceso de evaluación que, con base en un conjunto de criterios, indicadores y estándares, valida la calidad de la estructura, organización, funcionamiento, insumos, procesos de enseñanza, servicios y resultados de programas académicos y de instituciones de educación superior a través de una agencia u organismo acreditador.
Actitud	Disposición a actuar o pensar de determinada manera respecto a una situación o evento particular de acuerdo con nuestros valores, convicciones o intereses.
Alineación	Grado en que el contenido o las demandas cognitivas de las preguntas de la prueba se corresponden con el contenido o las demandas cognitivas objetivo.
Análisis de contenido	Método que permite identificar, inventariar y clasificar conceptos o ideas expresados en diferentes fuentes de información, a partir de una categorización establecida.
Anonimato	Condición que garantiza que la identidad y/o datos personales de los participantes involucrado en un proceso particular no serán divulgados bajo ninguna circunstancia.
Aprendizaje	Proceso mental primario para la adquisición de conocimientos, habilidades, comportamientos y actitudes posibilitado mediante el estudio, la enseñanza o la experiencia.
Aprendizaje a distancia	Situación en la que el alumno gestiona a su propio ritmo y horario las actividades de aprendizaje y su interacción con otros en diferentes tiempos y espacios físicos y/o virtuales, puede ser de manera sincrónica o asincrónica.
Asignatura	Contenido disciplinario contemplado en el mapa curricular de un plan de estudios que posee una clave oficial y tiene asociado un valor en créditos de acuerdo con la dedicación de horas de docencia o de trabajo total de los estudiantes.

TÉRMINO	DEFINICIÓN
Atributo	Cualidad o propiedad que caracteriza y da cierto significado a un sujeto u objeto.
Autoevaluación	Tipo de evaluación en la cual es el propio sujeto quien evalúa su desempeño de manera que le sea posible identificar aspectos que puede mejorar.
Banco de reactivos	Espacio físico o virtual que contiene reactivos con su información asociada (autor, fecha en que se dio de alta, número de veces que se ha aplicado, fecha de aplicación, etcétera.) así como sus características psicométricas.
Calibración	Análisis estadístico de los resultados de un instrumento de evaluación para determinar sus parámetros psicométricos (dificultad, discriminación, aporte a la calificación, ajuste al modelo teórico y pseudo adivinación) según la teoría de referencia que se utilice.
Calificación	Resultado del proceso de evaluación que refleja la cantidad del atributo que se está midiendo.
Certificación	Proceso mediante el cual se reconoce que las personas han demostrado un determinado nivel de conocimientos y capacidades en un dominio específico.
Coeficiente Alfa de Cronbach	Estadístico que permite estimar el grado de confiabilidad de un instrumento mediante el análisis de las variaciones de sus puntajes con respecto a la varianza total.
Coevaluación	Tipo de evaluación en la cual dos o más sujetos que se encuentran dentro de un mismo proceso formativo se evalúan entre sí.
Cohorte	Grupo de personas que comparten características similares entre sí o que experimentan una misma situación durante determinado lapso.

TÉRMINO	DEFINICIÓN
Competencia	Es la capacidad de resolver problemas mediante el uso y la movilización de recursos psicosociales incluyendo conocimientos, habilidades y actitudes en un contexto particular.
Confiabilidad	Grado en el que los resultados de un instrumento son consistentes en diferentes aplicaciones bajo las mismas condiciones.
Criterio	Se refiere a un valor que se establece y se define en un proceso para juzgar el mérito de un objeto o un componente.
Diseño de evaluación	Proceso a través del cual se adopta un conjunto de decisiones que justifican la elección de un enfoque, así como la de los procedimientos para realizar una evaluación.
Docencia	Acción de enseñar en un contexto de educación formal.
Educación	Proceso permanente para la adquisición, la transmisión y el acrecentamiento de ideas, conocimientos, habilidades, hábitos y valores por medio de la instrucción, el estudio, el ejemplo y la experiencia.
Educación a distancia	Modalidad de enseñanza-aprendizaje desarrollada principalmente de forma remota, se realiza cuando estudiantes y profesores se encuentran separados por la distancia física y la tecnología es empleada como un puente para reducir esta barrera.
Educación continua	Serie de actividades académicas extracurriculares que sirven para impulsar la formación y superación personal y profesional.
Eficacia	Logro de los propósitos o metas de un proceso, así como de los procedimientos utilizados para desempeñar determinadas funciones.
Eficiencia	Capacidad para lograr los objetivos propuestos con el mínimo esfuerzo y el menor costo posible, utilizando adecuadamente los recursos disponibles.

TÉRMINO	DEFINICIÓN
Eficiencia terminal	Se refiere al número total de alumnos que concluyen los requerimientos de un ciclo de estudios específico en el tiempo estipulado en su plan de estudios.
Egresados	Estudiantes que han cubierto 100% de los créditos establecidos en un programa académico en el tiempo determinado por las normas institucionales.
E-learning	Modalidad de enseñanza-aprendizaje que se lleva a cabo mediante el uso de Internet y tecnologías de la información y comunicación (TIC).
Enseñanza	Acción que favorece el entendimiento de conocimientos y el intercambio de experiencias a través del diseño, la selección y transmisión de contenido, la evaluación y la reflexión.
Estándar	Nivel, cualitativo y/o cuantitativo, de logro que se considera aceptable y que es previamente establecido por alguna agencia, organismo acreditador o institución.
Estrategias de enseñanza-aprendizaje	Procedimientos o recursos (organizadores del conocimiento) que son empleados por el docente a fin de promover el aprendizaje.
Evaluación	Emisión de juicios de valor como resultado de comparar una medición u observación con un referente previamente establecido.
Evaluación criterial	Tipo de evaluación en el que se valora el desempeño de un individuo contra un conjunto de criterios de aprendizaje, habilidades o atributos predeterminados, en donde la interpretación de los resultados de cada estudiante es independiente del nivel alcanzado por el grupo al cual pertenece.
Evaluación de alto impacto	Tipo de evaluación también conocida como evaluación de altas consecuencias debido a que los efectos que pueden tener sus resultados impactan no únicamente en la vida académica de los estudiantes, sino que pueden llegar más allá e incidir en otros ámbitos.

TÉRMINO	DEFINICIÓN
Evaluación de bajo impacto	Tipo de evaluación en la cual las consecuencias de los resultados no conllevan mayores problemas y generalmente se puede trabajar sobre las áreas de oportunidad que se identifiquen.
Evaluación de instituciones educativas	Campo especializado de la evaluación en el que se realizan procesos múltiples de valoración sobre los distintos componentes y procedimientos de una institución educativa.
Evaluación de la docencia	Campo especializado de la evaluación educativa en el que se valoran las características y el desempeño de los profesores, a través de distintos métodos.
Evaluación de un programa educativo	Proceso que permite identificar las áreas de oportunidad y fortalezas de un programa con el objetivo de emitir juicios de valor sobre el estado que guarda su quehacer académico y después llevar a cabo acciones que conduzcan a mejorar su calidad.
Evaluación del aprendizaje	Proceso de evaluación en el que se valoran los conocimientos y habilidades adquiridas por los estudiantes como resultado de diversas experiencias educativas.
Evaluación diagnóstica	Tipo de evaluación que tiene la finalidad de identificar los conocimientos o el nivel de desarrollo de las habilidades previas de los estudiantes.
Evaluación educativa	Disciplina en la cual se emiten juicios de valor sobre un objeto del ámbito educativo, con el fin de tomar decisiones particulares para su mejora.
Evaluación en línea	Uso de la tecnología de la información como medio para evaluar cuantitativa y/o cualitativamente el aprendizaje de los estudiantes.
Evaluación externa	Evaluación conducida por agentes externos a la institución o programa que se evalúa. Generalmente intervienen grupos de pares, comités de expertos u organismos especializados en cuestiones de evaluación.

TÉRMINO	DEFINICIÓN
Evaluación formativa	Tipo de evaluación que se centra en los procesos que se ponen en juego para el logro de los resultados de aprendizaje con la finalidad de detectar fortalezas y áreas de oportunidad en el aprendizaje de los alumnos.
Evaluación interna	Evaluación conducida por un miembro o miembros de la institución. Gran parte de las instituciones de educación superior emplean este tipo de evaluación.
Evaluación normativa	Tipo de evaluación en la cual por medio de los resultados se pueden realizar comparaciones del rendimiento de cada estudiante con respecto a un grupo y así determinar si tal alumno tiene un rendimiento superior o inferior al resto del grupo.
Evaluación para el aprendizaje	Evaluación en el cual los resultados se utilizan para identificar fortalezas y áreas de oportunidad en el proceso de aprendizaje de los estudiantes y mejoras en las técnicas de enseñanza empleadas por los profesores.
Evaluación sumativa	Tipo de evaluación en la cual se realiza un balance general de los conocimientos adquiridos o competencias desarrolladas después de un proceso de enseñanza.
Evaluador	Profesional responsable de planear y conducir procesos de evaluación en diferentes contextos y con diversos propósitos y alcances. Esta figura es desarrollada por los docentes.
Examen	Instrumento de evaluación que tiene como propósito medir el grado de dominio de conocimientos o aptitudes para valorar hasta qué punto es necesario mejorar los métodos o agentes a través de los cuáles éstos se adquieren.
Formación docente	Proceso orientado al desarrollo profesional y personal del profesorado que abarca una profundización de un marco teórico-conceptual sobre los procesos educativos que ocurren en el aula.
Habilidades	Características de una persona que indican sus capacidades físicas o mentales para desarrollar ciertas tareas.

TÉRMINO	DEFINICIÓN
Heteroevaluación	Tipo de evaluación en la cual quien evalúa es un agente externo al sujeto de evaluación.
Indicador	Valor cuantitativo o cualitativo que expresa las características o estado de un individuo, objeto o proceso.
Información cualitativa	Conjunto de hechos, percepciones, acciones o resultados expresados en forma descriptiva o narrativa.
Información cuantitativa	Conjunto de hechos, percepciones, acciones o resultados expresados numéricamente.
Informe de evaluación	Documento que contiene el diseño, la planeación, los resultados y recomendaciones de un proceso de evaluación.
Innovación educativa	Conjunto de intervenciones, decisiones y estrategias, intencionales y planeadas, que tienen como finalidad generar cambios significativos para la mejora de los procesos educativos.
Instrumento de evaluación	Técnica de medición y recolección de datos que tiene distintos formatos atendiendo a la naturaleza de la evaluación.
Juicio	Apreciación respecto al contenido, las características o desempeño de un determinado objeto que está siendo evaluado.
Lista de cotejo	Instrumento que contiene criterios o indicadores a partir de los cuales se miden y evalúan las características del objeto, comprobando si cumple con los atributos establecidos.
Mapa curricular	Representación gráfica de un plan, programa, módulo o asignatura educativa, que esquematiza organizadamente, la coherencia horizontal y vertical de los diferentes elementos que integran el plan de estudios.
Medición	Proceso mediante el cual se asignan numerales a características o atributos de un objeto o proceso a través de un conjunto de reglas definidas.

TÉRMINO	DEFINICIÓN
Muestra	Parte de una población a partir de la cual es posible hacer inferencias.
Objeto de evaluación	Componente que se evalúa, y sobre el cual se toman decisiones en función de un conjunto de criterios previamente establecidos.
Perfil de referencia	Documento que constituye un marco de referencia en el que se expresan las aptitudes, valores y conocimientos que se espera que posea el examinado.
Plan de estudios	Documento que, con fundamento en una orientación educativa, precisa y articula objetivos, contenidos y estrategias de enseñanza-aprendizaje para sustentar la formación de una persona en el contexto de un ciclo completo de educación.
Población	Conjunto total de individuos de un grupo que se sujeta a estudio o a evaluación de acuerdo con una o varias características o atributos comunes.
Proctoring	Metodología de supervisión de exámenes en la que se vigila de forma remota a los sustentantes, permitiendo que el examen se realice donde se encuentre el examinado y asegurando que el comportamiento sea el adecuado.
Programa de la asignatura	Documento que presenta un desglose temático, los objetivos de aprendizaje, los métodos de enseñanza y de aprendizaje, la bibliografía y los medios de evaluación de una asignatura.
Prueba piloto	Ensayo preliminar mediante el cual se ponen a prueba instrumentos o procesos con el fin de adecuar o ajustar sus características técnicas y aspectos de su aplicación.
Reactivo	Unidad de medida que consiste en una instrucción, pregunta, ejercicio, tarea o problema que requiere una respuesta del examinado, a partir de la cual se puede inferir su ejecución o desempeño en el objeto de la evaluación.

TÉRMINO	DEFINICIÓN
Retroalimentación	Comentarios, orales o escritos, acerca del desempeño académico en un proceso de evaluación o una actividad dentro de un escenario educativo, que permite mejorar los conocimientos, las habilidades y las actitudes para alcanzar un objetivo y desarrollar la capacidad de autorregulación.
Resultado de aprendizaje	Son descripciones de los conocimientos y habilidades que el examinado debe comprender y aplicar como producto de lo aprendido en su formación.
Tabla de especificaciones	Matriz que se emplea para diseñar un examen, incluye clave de temas y subtemas, temas y subtemas, resultados de aprendizaje y peso específico o ponderación.
Taxonomía de Bloom	Sistema de clasificación elaborado por Benjamin Bloom en el que presenta una jerarquía de seis niveles de las habilidades cognitivas que los seres humanos pueden desarrollar en el proceso de aprendizaje: (1) conocimiento, (2) comprensión, (3) aplicación, (4) análisis, (5) síntesis y (6) evaluación.
Trayectoria escolar	Recorrido que sigue una cohorte de estudiantes en un tiempo determinado, a partir de su ingreso a un plan de estudios o programa académico.
Validez	Juicio evaluativo e integrado, del grado al que la evidencia empírica y razonamientos teóricos apoyan lo adecuado y apropiado de las inferencias y acciones basadas en puntuaciones de exámenes y otras formas de evaluación.
Variable	Característica o propiedad de un objeto que tiende a cambiar al desarrollarse en ambientes o contextos diferentes y que puede tener diversos valores.

REFERENCIAS

AERA, APA, NCME. (2018). *Estándares para Pruebas Educativas y Psicológicas*. Recuperado el 31 de enero del 2022, de https://www.testingstandards.net/uploads/7/6/6/4/76643089/9780935302745_web.pdf

Asurza, H., y Alejandro, S. (2006). *Glosario básico de términos estadísticos*. INEI. LIMA: Centro de ediciones del INE. Recuperado el 3 de febrero del 2022, de https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib0900/Libro.pdf

Austrian Development Agency. (2009). *Guidelines for Project and Programme Evaluations*. Recuperado el 3 de febrero del 2022, de https://www.oecd.org/derec/austria/AUSTRIA%20ADA%20ADC%20Guidelines.pdf

Australian Government. (2020). *Norm and criterion referenced assessments*. Recuperado el 3 de febrero, de https://www.edcan.org.au/edcan-learning-resources/using-edcan-resources/implementation-resources/assessment/norm-criterion

Avanzo Learning Progress S.A. (2022, 31 de enero). ¿Qué es el e-*learning?* Avanzo. Recuperado el 3 de febrero del 2022, de https://www.avanzo.com/que-es-el-elearning/

Badillo, J. F. (2021, 11 de agosto). *Qué es e-learning, definición, características y ventajas*. TICAP. Recuperado el 3 de febrero del 2022, de https://www.ticap.mx/que-es-e-learning-definicion/

Bustos, N. (2020, 26 de mayo). *Enseñanza sincrónica y asincrónica ¿cuál es su diferencia?* BridgeEnglish. Recuperado el 3 de febrero del 2022, de https://bridgeenglish.cl/blog/ensenanza-sincronica-y-asincronica-cual-es-su-diferencia/

Casanova, M. A. (2007). Manual de evaluación educativa (9ª ed.). La Muralla.

Cambridge Assessment International Education. (2019). Evaluación para el aprendizaje. *UCLES* Recuperado el 3 de febrero del 2022, de https://www.cambridgeinternational.org/Images/579619-assessment-for-learning-spanish-.pdf

Choppin, B. (1976). *Developments in ítem banking*. En Monitoring National Standards of Attainment in Schools. Recuperado el 3 de febrero, de https://www.rasch.org/memo76.pdf

Clarke, M. & Luna-Bazaldua, D. (2021). Primer on Large-Scale Assessments of Educational Achievement. National Assessments of Educational Achievement series. Washington, DC: World Bank. Recuperado el 3 de febrero, de https://doi.org/10.1596/978-1-4648-1659-8

Consejo para la Acreditación de la Educación Superior A.C. (s. f.). ¿Qué es la Acreditación? Recuperado el 3 de febrero del 2022, de https://www.copaes.org/queesacreditacion.html

Delgado, P. (2020, 23 de junio). ¿Cuál es la diferencia entre aprendizaje sincrónico y asincrónico? *Observatorio de Innovación Educativa*. Recuperado el 3 de febrero del 2022, de https://observatorio.tec.mx/edu-news/aprendizaje-sincronico-y-asincronico-definicion

Dirección General de Planeación. (2018). *Glosario de términos estadísticos*. Recuperado el 3 de febrero del 2022, de https://www.planeacion.unam.mx/Agenda/2018/disco/xls/glosario-20181109.pdf

Dirección General de Planeación y Programación. (2008) *Glosario. Términos utilizados en la Dirección General de Planeación y Programación.* SEP Recuperado el 3 de febrero del 2022, de https://es.slideshare.net/DavidMrs/glosario-direccion-general-de-planeacion-sep

Docentes al día. (2019, 8 de diciembre). Coevaluación: qué es y ejemplo de instrumento para implementarla. *Docentes al día.* Recuperado el 3 de febrero del 2022, de https://docentesaldia.com/2019/12/08/coevaluacion-que-es-y-ejemplo-de-instrumento-para-implementarla/

Downing, S. M. (2004). Reliability on the reproducibility of assessment data. Med. Educ., Vol. 38:1006-1012. Recuperado el 31 de enero del 2022, de https://doi.org/10.1111/j.1365-2929.2004.01932.x

Educational Testing Service. (s.f). *Glossary of Standardized Testing Terms*. Recuperado el 3 de febrero del 2022, de https://www.ets.org/understanding_testing/glossary/

Educational, Scientific and Cultural Organization (s. f.). *Bloom´s taxonomy*. Recuperado el 3 de febrero del 2022, de http://www.ibe.unesco.org/es/node/12046

Fernández, F. (2018, 23 de febrero). La evaluación y su importancia en la educación. Nexos. Recuperado el 3 de febrero del 2022, de https://educacion.nexos.com.mx/?p=1016

Frechtling-Westat, J., Frierson, H., Hood, S., & Hughes, G. (2002). *The 2002 user friendly handbook for project evaluation.* Division of Research, Evaluation and Communication National Science Foundation. Recuperado el 3 de febrero del 2022, de https://www.nsf.gov/pubs/2002/nsf02057/nsf02057.pdf

Ganduxé, M. (2020, octubre). ¿Qué es el e-learning? *eLearning Actual*. Recuperado el 3 de febrero del 2022, de https://elearningactual.com/e-learning-significado/

Great Schools Partnership. (2014, 5 de marzo). Bloom´s taxonomy. *The glossary of Education Reform*. Recuperado el 3 de febrero del 2022, de https://www.edglossary.org/blooms-taxonomy/

Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research.* Recuperado el 3 de febrero del 2022, de https://doi.org/10.3102/003465430298487

Houghton Mifflin Harcourt. (s. f.). *Glossary of Testing, measurement, and statistical terms.* Recuperado el 3 de febrero del 2022, de https://www.hmhco.com/~/media/sites/home/hmh-assessments/hmhassessments_testingtermglossary.pdf?la=en

IGI Global Publisher of Timely knowledge. (s. f.). What is *Teaching*. Recuperado el 3 de febrero del 2022, de https://www.igi-global.com/dictionary/broadcasting-transforming-social-construction-knowledge/29344

IGI Global Publisher of Timely knowledge. (s. f.). *What is Teaching-Learning Process*. Recuperado el 3 de febrero del 2022, de https://www.igi-global.com/dictionary/teaching-learning-process/48941

IGI Global Publisher of Timely knowledge. (s. f.). *What is E-Assessment?* Recuperado el 3 de febrero del 2022, de https://www.igi-global.com/dictionary/tools-for-e-assessment-techniques-in-education/8580

Instituto Nacional para la Evaluación de la Educación. (2019). *Cuadernillo técnico de evaluación educativa. Definición del referente de la evaluación y desarrollo del marco de especificaciones*. Recuperado el 3 de febrero del 2022, de https://historico.mejoredu.gob.mx/wp-content/uploads/2019/08/P2A353.pdf

Instituto Nacional para la Evaluación de la Educación. (2015). *Política Nacional de la Evaluación de la Educación.* Recuperado el 31 de enero del 2022, de https://www.inee.edu.mx/wp-content/uploads/2019/01/P1E302.pdf

Joint United Nations Programme on HIV/AIDS. (s. f.). *Monitoring and Evaluation Terms*. Recuperado el 3 de febrero del 2022, de https://www.unaids.org/sites/default/files/sub_landing/files/11_ME_Glossary_FinalWorkingDraft.pdf

Khan, Z. (s. f.). Diagnostic Evaluation and Remedial Instruction. *Your article library.* Recuperado el 3 de febrero del 2022, de https://www.yourarticlelibrary.com/education/diagnostic-evaluation-and-remedial-instruction/45222

Leyva Barajas, Y. E. (2010). Evaluación del Aprendizaje: Una guía práctica para profesores. *Universidad Nacional Autónoma de México.* Recuperado el 3 de febrero del 2022, de https://www.ses.unam.mx/curso2012/pdf/Guia_evaluacion_aprendizaje2010.pdf

Lusquiños, C. (2016) La formación docente en la escuela. *Aprender a enseñar. El desafío de la formación docente inicial y continua.* Recuperado el 3 de febrero del 2022, de https://www.teseopress.com/formaciondocente/chapter/170/

Malamed, C. (2019). Glossary of Online Learning Terms. T*he eLearning Coach. Helping your design smarter learning experiences.* Recuperado el 3 de febrero del 2022, de https://theelearningcoach.com/resources/online-learning-glossary-of-terms/

Martínez, F. J., García, M., Infante Moro, A., Gallardo, J., & Infante Moro, J. C. (2018). *Proctoring: reto para la enseñanza del siglo XXI.* ResearchGate. Recuperado el 3 de febrero del 2022, de https://www.researchgate.net/publication/324676579_Proctoring_reto_para_la_ensenanza_del_siglo_XXI

Mendoza, Y. L y Mamani, J. E. (2012). Estrategias de enseñanza-aprendizaje de los docentes de la Facultad de Ciencias Sociales de la Universidad Nacional del Altiplano- Puno 2012. *Revista de Investigación en Comunicación y Desarrollo*, *3*(1), 58-67. Recuperado el 3 de febrero del 2022, de https://www.redalyc.org/pdf/4498/449845035006.pdf

Messick, S. (1989). Validity. In: R. L. Linn (Ed.), Educational measurement. 3rd ed. 13–103. New York: American Council on Education/Macmillan.

Michael, T. B., & Williams, M. A. (2013). Student equity: Discouraging cheating in online courses. *Administrative Issues Journal*, *3*(2), 6. Recuperado el 3 de febrero del 2022, de https://www.researchgate.net/publication/259150822_Student_Equity_Discouraging_Cheating_in_Online_Courses

Mimenza, O. C., y Castillero Mimenza, O. (2020, 21 de noviembre). *Diccionario de Psicología: 200 conceptos fundamentales.* Psicología y Mente. Recuperado el 3 de febrero del 2022, de https://psicologiaymente.com/psicologia/diccionario-psicologia

Morán Oviedo, P. (2004). La docencia como recreación y construcción del conocimiento Sentido pedagógico de la investigación en el aula. *Perfiles educativos*, *26*(105-106), 41-72. Recuperado el 3 de febrero del 2022, de http://www.scielo.org.mx/scielo.php?pid=S0185-26982004000100003&script=sci_abstract&tlng=es

Murillo, A. (2017, 3 de octubre). ¿Qué es la innovación educativa? *Observatorio de Innovación Educativa.* Recuperado el 3 de febrero del 2022, de https://observatorio.tec.mx/edu-news/innovacion-educativa

Nacional Center for Education Statistics. (2021). *Projections of Education Statistics to 2021 - Glossary*. Recuperado el 3 de febrero de 2022, de https://nces.ed.gov/programs/projections/projections2021/glossary.asp

National Council on Measurement in Education (s. f.). *Assessment glossary*. Recuperado el 3 de febrero del 2022, de https://www.ncme.org/resources/glossary

Nissan, S. & French, R. (2013). Item Banking. *Wiley Online Library*, *2*(7). Recuperado el 3 de febrero del 2022, de https://doi.org/10.1002/9781118411360.wbcla075

Organisation for Economic Cooperation and Development. (2010). *Glossary of Key Terms in Evaluation and Result Based Management*. Publicaciones OCDE: Francia. Recuperado el 3 de febrero del 2022, de https://www.oecd.org/dac/evaluation/2754804.pdf

Organisation for Economic Cooperation and Development. (2005). *The definition and selection of key competencies*. *Executive Summary*. Recuperado el 3 de febrero del 2022, de http://www.oecd.org/pisa/35070367.pdf

Oxford Reference (s. f.). *Bloom´s taxonomy*. Recuperado el 3 de febrero del 2022, de https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095513202

Pérez, J., y Gardey, A. (2021). Definición de aprendizaje. Definición. Recuperado el 31 de enero del 2022, de https://definicion.de/aprendizaje/

Picardo, O., y Escobar, J. C. (2004). Diccionario Pedagógico UPAEP. *Centro de Investigación Educativa, San Salvador.* Recuperado el 3 de febrero del 2022, de https://otrasvoceseneducacion.org/wp-content/uploads/2019/02/diccionario.pdf

Priscila, (2021, 10 de septiembre). Aprendizaje sincrónico vs asincrónico. *EasyLMS*. Recuperado el 3 de febrero del 2022, de https://www.easy-lms.com/es/centro-de-conocimiento/centro-de-conocimiento-lms/aprendizaje-sincronico-vs-asincronico/item10387#useful-resources

Puerta, A. R. (2019, 13 de julio). *Heteroevaluación: características, objetivos, ventajas, desventajas*. Lifeder. Recuperado el 3 de febrero del 2022, de https://www.lifeder.com/heteroevaluacion/

Real Academia Española. (2014). *Diccionario de la lengua española*. Recuperado el 3 de febrero del 2022, de https://www.rae.es/

Red Iberoamericana para la Acreditación de la Calidad de la Educación Superior (2004). *Glosario Internacional RIACES de Evaluación de la Calidad y Acreditación*. Recuperado el 3 de febrero del 2022, de http://riaces.org/glosario/

Riquelme, M. (2022, 6 de enero). Estrategias de Aprendizaje (definición y clasificación). *Web y empresas*. Recuperado el 3 de febrero del 2022, de https://www.webyempresas.com/estrategias-de-aprendizaje/

Rodríguez, F., Ibarra, M. S. y García, E. (2013). Autoevaluación, evaluación entre pares y coevaluación: conceptualización y práctica en las universidades españolas. *Revista de Investigación en Educación*, *11*(2), 198-210. Recuperado el 3 de febrero del 2022, de https://idus.us.es/bitstream/handle/11441/68957/708-1788-1-PB.pdf;jsessionid=AD4F82E22779E0473D78D01AE325CD4C?sequence=1&isAllowed=y

Sawchuk, S. (2015, 3 de septiembre). Teacher Evaluation: An Issue Overview. *Education Week*. Recuperado el 3 de febrero del 2022, de https://www.edweek.org/ew/section/multimedia/teacher-performance-evaluation-issue-overview.html

Service Efforts and Accomplishments. (s. f.). *Glossary*. Recuperado el 3 de febrero del 2022, de http://www.seagov.org/resources/

Starr-Glass, D. (2019). *Engagement, Publishing, and the Scholarship of Teaching and Learning: Reconsidering the Reconsidered*. Recuperado el 3 de febrero del 2022, de https://www.igi-global.com/chapter/engagement-publishing-and-the-scholarship-of-teaching-and-learning/217555

Teaching and Learning in Higher Education. (s. f.). *Assessment Types: Diagnostic, Formative and Summative*. Recuperado el 3 de febrero del 2022, de https://www.queensu.ca/teachingandlearning/modules/assessments/10_s2_02_diagnostic_assessment.html

United States Agency for International Development. (2009). *Glossary of evaluation terms*. Recuperado el 3 de febrero del 2022, de https://pdf.usaid.gov/pdf_docs/Pnado820.pdf

United States Environmental Protection Agency (2022, enero 18). *Program Evaluation Glossary*. Recuperado el 3 de febrero del 2022, de https://ofmpub.epa.gov/sor_internet/registry/termreg/searchandretrieve/glossariesandkeywordlists/search.do?details=&glossaryName=Program%20Evaluation%20Glossary

University of Massachusetts Amherst. (s. f.). *Criterion and Norm-Based Evaluation*. Recuperado el 3 de febrero del 2022, de https://www.umass.edu/oapa/srti/srti-and-performance-appraisal/criterion-and-norm-based-evaluation

Universidad de Murcia. (s. f.). *Glosario de términos.* Recuperado el 3 de febrero del 2022, de https://www.um.es/c/document_library/get_file?uuid=7613b273-d972-445a-872f-9b0d63aaf122&groupId=1277604

Universidad de Valencia. (2010). SPSS: Análisis de fiabilidad. Alpha de Cronbach. Grupo de innovación educativa. Recuperado el 31 de enero del 2022, de https://www.uv.es/innomide/spss/SPSS/SPSS_0801B.pdf

Universidad Estatal de Distancia de Costa Rica. (s. f.). *Glosario de términos curriculares para la Universidad Estatal a Distancia.* Recuperado el 3 de febrero del 2022, de https://www.uned.ac.cr/academica/images/PACE/publicaciones/FINAL24-9-13_Glosario_de_trminos_curriculares_UNED.pdf

Universidad Nacional Autónoma de México. (2008). *Glosario de Educación a distancia.* Recuperado el 3 de febrero del 2022, de https://www2.uned.es/catedraunesco-ead/Colaboraciones/varios/Glosario.pdf

Universidad Nacional Autónoma de México. (2013). *Glosario de Términos estadísticos UNAM* [Archivo PDF]. Recuperado el 3 de febrero del 2022, de http://agendas.planeacion.unam.mx/pdf/Glosario.pdf

Universidad Nacional Autónoma de México. (s. f.). *Perfil de referencia* [Video]. Coursera. Recuperado el 3 de febrero del 2022, de https://es.coursera.org/lecture/evaluacion-aprendizaje-enfoque-cuantitativo/perfil-de-referencia-de4Jh

Universiteit Iyunivesithi Stellenbosch University. (s.f.). *Teaching, learning, assessment, curriculum and pedagogy*. Recuperado el 3 de febrero del 2022, de http://www.sun.ac.za/english/learning-teaching/ctl/t-l-resources/curriculum-t-l-assessment

University of Texas. (2012). *What is competency?* Recuperado el 3 de febrero del 2022, de https://sph.uth.edu/content/uploads/2012/01/Competencies-and-Learning-Objectives.pdf

Weleschuk, A., Dyjur, P. & Kelly, P. (2019). Online Assessment in Higher Education. *Taylor Institute for Teaching and Learning Guide Series. Calgary, AG: Taylor Institute for Teaching and Learning at the University of Calgary.* Recuperado el 3 de febrero del 2022, de https://taylorinstitute.ucalgary.ca/sites/default/files/TI%20Guides/Online%20Assessment%20Guide-2019-10-24.pdf

AUTORES

Adrián Alejandro Martínez González
Director de Evaluación Educativa de la Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia (CUAIEED), UNAM. Profesor de Carrera Titular "C" Tiempo Completo Definitivo de la Facultad de Medicina, UNAM.
Correo electrónico: adrian_martinez@cuaieed.unam.mx
ORCID: https://ORCID.org/0000-0002-5021-9639

Abigail Patricia Manzano Patiño
Jefa del Departamento de Análisis de Resultados de Evaluación Formativa y de Diagnóstico de la CUAIEED, UNAM.
Correo electrónico: abigail_manzano@cuaieed.unam.mx
ORCID: https://ORCID.org/0000-0003-0434-0654

Adrián Israel Martínez Franco
Associate Professor and Department Head of Rural Community Development and Community Health, Abraham Baldwin Agricultural College.
Correo electrónico: amartinezfranco@abac.edu
ORCID: https://ORCID.org/0000-0002-4631-5425

Alan Kristian Hernández Romo
Coordinador de Gestión de la CUAIEED, UNAM. Profesor de la Facultad de Medicina Veterinaria y Zootecnia.
Correo electrónico: alan_hernandez@cuaieed.unam.mx
ORCID: https://ORCID.org/0000-0002-4299-8628

Alex J. Mechaber
Vicepresidente del Examen de Licencia Médica de los Estados Unidos (United States Medical Licensing Examination, USMLE) del National Board of Medical Examiners (NBME), EUA
Correo electrónico: amechaber@nbme.org
ORCID: https://ORCID.org/0000-0002-0277-8428

Amanda L. Clauser
Vicepresidenta de Psicometría y Análisis de Datos del National Board of Medical Examiners (NBME), EUA.
Correo electrónico: claruser@nbme.org
ORCID: https://ORCID.org/0000-0002-1202-3862

Ana Itzel Pascual Vigil
Asesora en Evaluación de la Subdirección de Evaluación de Posgrado y Titulación de la Dirección de Evaluación Educativa de la CUAIEED, UNAM.
Correo electrónico: a.pascual.vigil@gmail.com
ORCID: https://ORCID.org/0000-0003-2397-7087

Ana Laura Pérez Díaz
Coordinadora de Evaluación de Posgrado y Titulación de la Dirección de Evaluación Educativa de la CUAIEED, UNAM.
Correo electrónico: ana_perez@cuaieed.unam.mx

Angélica Ramírez Elías
Profesora de la Escuela Nacional de Enfermería y Obstetricia, UNAM.
Correo electrónico: arelias1999@yahoo.com
ORCID: https://ORCID.org/0000-0002-8015-5840

Anisai Ledesma Rodea
Estudiante del doctorado en la Facultad de Psicología, UNAM.
Correo electrónico: anisai.758@gmail.com
ORCID: https://ORCID.org/0000-0002-0662-897X

Benilde García Cabrero
Profesora en los Programas de Maestría y Doctorado del Campo de Psicología Educativa y del Desarrollo y de la Maestría en Docencia en Educación Media Superior en la División de Investigación y Posgrado de la Facultad de Psicología, UNAM.
Correo electrónico: benildegar@gmail.com
ORCID: https://ORCID.org/0000-0002-8208-5284

Blanca Ariadna Carrillo Ávalos
Profesora Investigadora de Tiempo Completo de la Facultad de Medicina de la Universidad Autónoma de San Luis Potosí.
Correo electrónico: ariadna.carrillo@uaslp.mx
ORCID: https://ORCID.org/0000-0003-4111-4795

Careli Johana Herrera Penilla
Jefa de la Unidad Académica de Trayectorias Escolares de la Dirección de Evaluación Educativa de la CUAIEED, UNAM.
Correo electrónico: pedagogiaffyl@gmail.com
ORCID: https://ORCID.org/0000-0002-6947-7996

Carlos Gutiérrez-Cirlos Madrid
Profesor del Pregrado, Coordinador de los cursos de la Facultad de Medicina de la UNAM y Médico Especialista del Instituto Nacional de Ciencias Médicas y Nutrición "Salvador Zubirán".
Correo electrónico: cirlos@hotmail.com
ORCID: https://ORCID.org/0000-0001-9260-5370

Cinthya Trejo Rojas
Líder de Proyecto de la Subdirección de Evaluación de Posgrado y Titulación de la Dirección de Evaluación Educativa de la CUAIEED, UNAM.
Correo electrónico: cinthya_trejo@cuaieed.unam.mx

David A. Johnson
Funcionario Principal de Evaluaciones de la Federación de Consejos Médicos Estatales (Federation of State Medical Boards).
Correo electrónico: djohnson@fsmb.org
ORCID: https://ORCID.org/0000-0003-3669-1838

Edna Luna Serrano
Instituto de Investigación y Desarrollo Educativo (IIDE) de la Universidad Autónoma de Baja California (UABC).
Correo electrónico: eluna@uabc.edu.mx
ORCID: http://ORCID.org/0000-0003-1496-548X

Eduardo Backhoff Escudero
Presidente de la Junta de Gobierno del INEE y de Métrica Educativa, A.C.
Correo electrónico: ebackhoff@gmail.com
ORCID: https://ORCID.org/0000-0001-7267-4774

Elibidú Ortega Sánchez
Jefa del Departamento de Exámenes de Selección de Bachillerato de la Dirección de Evaluación Educativa de la CUAIEED, UNAM.
Correo electrónico: elibidu_ortega@cuaieed.unam.mx
ORCID: https://ORCID.org/0000-0003-1357-8795

Enrique Ricardo Buzo Casanova
Subdirector de Evaluación de Bachillerato y Licenciatura de la CUAIEED, UNAM.
Correo electrónico: enrique_buzo@cuaieed.unam.mx
ORCID: https://ORCID.org/0000-0001-7490-7826

Gabriela de la Cruz Flores
Profesora de Asignatura en la Facultad de Psicología e Investigadora de tiempo completo del Instituto de Investigaciones sobre la Universidad y la Educación, UNAM.
Correo electrónico: gabydc74@yahoo.com.mx
ORCID: https://ORCID.org/0000-0001-8997-1636

Gabriela González Lerma
Escuela de Medicina y Ciencias de la Salud del Tecnológico de Monterrey.
Correo electrónico: dra.gabrielaglzl@gmail.com
ORCID: https://ORCID.org/0000-0002-1635-7771

Guadalupe Soto Estrada
Coordinadora de Evaluación de Competencias y Desempeño de la Dirección de Evaluación Educativa de la CUAIEED, UNAM.
Correo electrónico: guadalupe_soto@cuaieed.unam.mx
ORCID: https://ORCID.org/0000-0003-1343-2016

Irma Elisa Eraña-Rojas
Profesora Asociada y Directora de Carrera del Programa de Médico Cirujano en Campus Monterrey del Tecnológico de Monterrey.
Correo electrónico: ierana@tec.mx
ORCID: https://ORCID.org/0000-0002-9022-3739

Isabel Hilliger Carrasco
Profesora Asistente de la Planta Especial y Subdirectora de Medición, Evaluación y Calidad de la Escuela de Ingeniería de la Pontificia Universidad Católica de Chile (PUC-Chile).
Correo electrónico: ihillige@ing.puc.cl
ORCID: http://ORCID.org/0000-0001-5270-7655

Iwin Leenen
Investigador de la Facultad de Psicología de la UNAM.
Correo electrónico: iwin.leenen@gmail.com
ORCID: https://ORCID.org/0000-0003-4807-540X

José de Jesús Naveja Romero
Doctor en Ciencias Egresado del Plan de Estudios Combinados en Medicina (PECEM).
Especialista en aplicación de modelos estadísticos.
Correo electrónico: navejaromero@gmail.com
ORCID: https://ORCID.org/0000-0001-8640-6690

José Gerardo Moreno Salinas
Subdirector de Proyectos de Transformación Digital para la Educación de la CUAIEED, UNAM.
Correo electrónico: gerardo_moreno@cuaieed.unam.mx
ORCID: https://ORCID.org/0000-0001-6539-0749

Karla Patricia Sosa Ramírez
Asesora en la evaluación de la docencia y tutoría, de planes y programas de estudios, estudios con egresados y procesos de acreditación.
Correo electrónico: karlasr.unam@gmail.com

Katherina Edith Gallardo Córdova
Directora del Doctorado en Innovación Educativa de la Escuela de Humanidades y Educación del Tecnológico de Monterrey.
Correo electrónico: katherina.gallardo@tec.mx
ORCID: https://ORCID.org/0000-0001-8343-9518

Laura Delgado Maldonado
Ex-Directora General de Medición y Tratamiento de Datos en el INEE y Ex-Directora de Área de Diseño, Ingenierías y Arquitectura en el CENEVAL
Correo electrónico: lau.delma@gmail.com

Laura Elena Rojo Chávez
Subdirectora de Evaluación de Procesos y Programas Académicos de la CUAIEED, UNAM.
Correo electrónico: laura_rojo@cuaieed.unam.mx

Laura Silvia Hernández Gutiérrez
Departamento de Integración de Ciencias Médicas. Facultad de Medicina UNAM. Instructor Avanzado Nivel 2 en Simulación Clínica Certificado por EUSim.
Correo electrónico: lhernandezmd@facmed.unam.mx
ORCID: https://ORCID.org/0000-0001-5705-207X

Manuel García Minjares
Coordinador de Análisis de Resultados de Evaluación Educativa de la CUAIEED, UNAM.
Correo electrónico: manuel_garcia@cuaieed.unam.mx
ORCID: https://ORCID.org/0000-0002-9535-5917

María Azucena Montoya Magno
Jefa de Departamento de la Subdirección de Evaluación de Posgrado y Titulación de la Dirección de Evaluación Educativa de la CUAIEED, UNAM.
Correo electrónico: azucena_montoya@cuaieed.unam.mx

María del Pilar Rico Sánchez
Doctorante en Pedagogía, Facultad de Filosofía y Letras de la UNAM.
Correo electrónico: mariarico@filos.unam.mx

María Elena Pérez Rivera
Prestadora de servicios profesionales en la Subdirección de Evaluación al Posgrado y Titulación de la Dirección de Evaluación Educativa de la CUAIEED, UNAM.
Correo electrónico: elenaprzr@gmail.com

María Juliana Londoño Cárdenas
Jefa de Departamento de la Dirección de Evaluación Educativa de la CUAIEED, UNAM.
Correo electrónico: juliana_londono@cuaieed.unam.mx
ORCID: https://ORCID.org/0000-0003-0146-3958

Mario Rueda Beltrán
Investigador Titular del Instituto de Investigaciones sobre la Universidad y la Educación (IISUE) de la UNAM.
Correo electrónico: mariorb@unam.mx
ORCID: http://ORCID.org/0000-0001-5004-4915

Melchor Sánchez Mendiola
Coordinador de la Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia (CUAIEED), UNAM. Profesor de Carrera Titular "C" de Tiempo Completo Definitivo, División de Estudios de Posgrado, Facultad de Medicina, UNAM.
Correo electrónico: melchor_sanchez@cuaieed.unam.mx
ORCID: https://ORCID.org/0000-0002-9664-3208

Miguel A. Paniagua
Vicepresidente Asociado de Operaciones de Evaluación y Educación Médica del Consejo Nacional de Examinadores Médicos (National Board of Medical Examiners, NBME).
Correo electrónico: mpaniagua@nbme.org
ORCID: https://ORCID.org/0000-0003-2307-4873

Mildred Vanessa López Cabrera
Decana Asociada Académica de la Escuela de Medicina y Ciencias de la Salud del Tecnológico de Monterrey.
Correo electrónico: mildredlopez@tec.mx
ORCID: https://ORCID.org/0000-0002-6965-6636

Nancy Fabiola Pérez Herrera
Profesora de Asignatura en la Maestría en Educación en Ciencias de la Salud de la Facultad de Medicina, UNAM.
Correo electrónico: perez.fa13@gmail.com
ORCID: https://ORCID.org/0000-0002-2008-3216

Nancy Sofía Contreras Michel
Subdirectora de Evaluación de Posgrado y Titulación en la Dirección de Evaluación Educativa de la CUAIEED, UNAM.
Correo electrónico: nancy_contreras@cuaieed.unam.mx
ORCID: https://ORCID.org/0000-0003-1557-961X

Olivia Espinosa Vázquez
Profesora de Carrera Titular Tiempo Completo en la Facultad de Odontología y Tutora de MADEMS, UNAM.
Correo electrónico: oliviaeducisal@fo.odonto.unam.mx
ORCID: https://ORCID.org/0000-0002-7323-003X

Rafael Enrico Sánchez Mayorga
Especialista en Políticas Públicas en Evaluación Educativa con experiencia en el diseño y elaboración de instrumentos y análisis psicométricos.
Correo electrónico: enricosanz@gmail.com

Roberto Santos Solórzano
Jefe de Departamento de Innovación Educativa de la CUAIEED, UNAM.
Correo electrónico: roberto_santos@cuaieed.unam.mx
ORCID: https://ORCID.org/0000-0002-2469-0048

Roxana Sierra González
Colabora en la Dirección de Evaluación Educativa de la CUAIEED, UNAM.
Correo electrónico: roxana_sierra@cuaieed.unam.mx

Sandra Paola Rodríguez Castillo
Jefa de área de la Dirección de Evaluación Educativa, CUAIEED, UNAM.
Correo electrónico: sandra_rodriguez@cuaieed.unam.mx
ORCID: https://ORCID.org/0000-0002-4409-1739

Silvia Lizett Olivares Olivares
Directora de Posgrados del Tecnológico de Monterrey.
Correo electrónico: solivares@tec.mx
ORCID: https://ORCID.org/0000-0002-8311-9589

Tiburcio Moreno Olivos
Departamento de Tecnologías de la Información de la División de Ciencias de la Comunicación y Diseño de la Universidad Autónoma Metropolitana-Cuajimalpa.
Correo electrónico: tmoreno@cua.uam.mx
ORCID: https://ORCID.org/0000-0002-0392-6621

Verónica Luna de Luz
Subdirectora de Innovación Educativa de la Dirección de Innovación Educativa, Desarrollo Curricular y Formación Docente de la CUAIEED, UNAM.
Correo electrónico: veronica_luna@cuaieed.unam.mx
ORCID: https://ORCID.org/0000-0001-5875-1681

Virginia González Garibay
Asesora en la evaluación de planes de estudio y de programas, en la evaluación de la docencia y de la tutoría, así como en procesos de acreditación y estudios con egresados.
Correo electrónico: virginia_gonzalez@cuaieed.unam.mx

Evaluación y aprendizaje en educación universitaria:
estrategias e instrumentos

Se terminó de editar e imprimir en el mes de junio de 2022, por Imagia Comunicación,
Tejocotes 56-7, Col. Tlacoquemecatl. Benito Juárez, 03200, CDMX.
Para su composición se utilizaron tipos Minion Pro y Roboto,
para los títulos, para figuras y tablas, Myriad Pro.
Se imprimieron 1000 ejemplares.

La educación universitaria en la era moderna enfrenta múltiples retos para satisfacer las demandas formativas de los estudiantes del siglo XXI. Uno de los principales retos, que a la vez constituye un área de oportunidad, es la evaluación del, para y como aprendizaje, con un amplio abanico de perspectivas, estrategias e instrumentos. La evaluación no debe ser un hecho aislado, sino un conjunto de tareas que se retroalimentan mutuamente con las actividades de enseñanza y aprendizaje. Por tanto, la evaluación educativa debe dar pie a realizar ajustes en las actividades de enseñanza y aprendizaje durante las diferentes fases del proceso formativo, que informen la toma de decisiones de docentes e instituciones educativas, al tiempo que propicien el aprendizaje profundo del estudiantado.

Los capítulos desarrollados en este libro pretenden ser una fuente de información, conocimientos, estrategias e instrumentos sobre diversos temas de evaluación y aprendizaje, para personal docente y funcionariado universitario de cualquier área del conocimiento, así como para estudiantes interesados en el tema. Para ello se describen y analizan aspectos conceptuales, metodológicos y prácticos sobre la evaluación del, para y como aprendizaje, en escenarios universitarios. Así mismo, se proporciona información y conocimientos útiles, relevantes y prácticos para el profesorado que desee enriquecer sus habilidades como evaluador en el dinámico mundo actual.

Este libro da respuesta a las necesidades que, en el ámbito de la evaluación educativa, fueron identificadas por el profesorado universitario y el área de evaluación educativa de la Universidad Nacional Autónoma de México (UNAM), en el ejercicio de sus tareas y después de una amplia revisión de la literatura sobre el tema. Constituye la suma de conocimientos y experiencias del personal del área de Evaluación Educativa de la Coordinación de Universidad Abierta, Innovación Educativa y Educación a Distancia (CUAIEED) de nuestra institución, así como de académicos y expertos en evaluación de diversas instituciones públicas y privadas que, con iniciativa y creatividad, han desarrollado prácticas de evaluación exitosas y útiles en una gran variedad de contextos.






9 786073 060769